

STIC-Biotech/ChemLib

From: Seharaseyon, Jegatheesan
Sent: Thursday, July 12, 2001 10:37 AM
To: STIC-Biotech/ChemLib
Subject: 09/554,933

Please search SEQ ID No: 3 of 09/554,933. Thanks.

J.Seharaseyon
Art Unit 1647
CM1 10D16
10CO1 MB
(703)-305-1112

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

PT Novel proteins containing transmembrane domains, useful as
 PT anti-inflammatory, immune stimulators/suppressors and tissue
 growth compounds

PS Claim 1: Page 68-69; 89pp; English.

CC This invention describes novel human transmembrane containing proteins
 CC and their encoding nucleic acids. Although no specific use is given for
 CC the proteins, they may have a range of activities selected from
 CC nutritional uses, cytokine and cell differentiation, immune
 CC stimulation/suppression, haematopoiesis regulation, tissue growth,
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined
 CC as gene sources for gene therapy. The cDNAs can also be used for large
 CC scale expression of proteins. The cDNAs can also be used for large
 CC detection of the corresponding ligands and for screening of novel
 CC low-molecular pharmaceuticals.

CC Sequence 383 AA;

Query Match

Best Local Similarity 100.0%; Score 2080; DB 20; Length 383;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGTGLFLFLFLLCAVGOVSPYSAPWKPMPAYRLPVLPSTINIAKPPDGAEAKE 60
 DB 1 magtglflflflfllcavgvspysapwkpmpayrlpvlpstlniakpdtgaekle 60
 QY VSSSCGPGCHGTPPLPYEAKOYLSEYELIYANGSRTETOVGIYILSSSGDAQHSDGS 120
 DB 61 vssscgpgchktplpyeakqylseyelyangsrtetqvgiylsssgdaqhrdsqs 120
 QY 121 SGSRRRKQIYGYDSRFSIFGKDFLNTPEFSTSVKLSGCTGTVAEKHVLTAAHCIDG 180
 DB 121 sgksrrkrqiygydsrfsifgkdfllnypfstsvklsctgtvaekhvltaa hcidg 180
 QY 181 KTYVKTOKLRVGFLLKPKKRCGGRANDSTAMPQMKFOWIRVKTVPVPGWIKGNAND 240
 DB 181 ktyvktqrlrvgflfkpkktdggrandsampsqmkfgyrlrvkthvpgwtkgnand 240
 QY 241 IGMDDYALIELLKKPKRRKFMKIGVSPPAKQIPGRIHFSGYDNDRCNLYRRCVYKDE 300
 DB 241 igmdyalyalllkkpkrrkfmki gvspakqipggrihfsydnndrcnlyr rcdvkd 300
 QY 301 TYDLTYOCCDAOPGASGSGVYVYRMKROQKWERKIIIGFSGHQWVDMNGSPDDFNVAVR 360
 DB 301 tydllyqccdaqpgasgsqgyvyrvmkrtqqkwerkiiigifsgbqwdmngspddfnvavr 360
 QY 361 ITPLKYAQICWIKGNVLDCREG 383
 DB 361 itplkyaqicwyikgnvldcreg 383

RESULT 2

ID AAY08660 standard; Protein; 383 AA.

AC AAY08660;

DE 09-AUG-1999 (first entry)

XX WO9927094 Seq ID 12.

KW Transmembrane domain; human; nutrition; cytokine; cell differentiation;
 KW immune stimulation; immune suppression; haematopoiesis; activin;
 KW regulatory tissue growth; inhibin; chemostatic; chemokinetic;
 KW haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;
 KW gene therapy; screening.

OS Homo sapiens.

XX

PN WO9927094-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-JP05238.

XX 25-NOV-1997; 97JP-0323129.

XX (PROT-) PROTEGENE INC.

XX (SAGA) SAGAMI CHEM RES CENT.

XX Kato S, Kimura T, Sekine S;

XX WPI; 1999-357835/30.

PT Novel proteins containing transmembrane domains, useful as
 PT anti-inflammatory, immune stimulators/suppressors and tissue
 growth compounds

PS Disclosure; Page 87-89; 89pp; English.

CC This invention describes novel human transmembrane containing proteins
 CC and their encoding nucleic acids. Although no specific use is given for
 CC the proteins, they may have a range of activities selected from
 CC nutritional uses, cytokine and cell differentiation, immune
 CC stimulation/suppression, haematopoiesis regulation, tissue growth,
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined
 CC as gene sources for gene therapy. The cDNAs can also be used for large
 CC scale expression of proteins. The cDNAs can also be used for large
 CC detection of the corresponding ligands and for screening of novel
 CC low-molecular pharmaceuticals.

SO Sequence 383 AA;

Query Match

Best Local Similarity 100.0%; Score 2080; DB 20; Length 383;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGTGLFLFLFLLCAVGOVSPYSAPWKPMPAYRLPVLPSTINIAKPPDGAEAKE 60
 DB 1 magtglflflflfllcavgvspysapwkpmpayrlpvlpstlniakpdtgaekle 60
 QY VSSSCGPGCHGTPPLPYEAKOYLSEYELIYANGSRTETOVGIYILSSSGDAQHSDGS 120
 DB 61 vssscgpgchktplpyeakqylseyelyangsrtetqvgiylsssgdaqhrdsqs 120
 QY 121 SGSRRRKQIYGYDSRFSIFGKDFLNTPEFSTSVKLSGCTGTVAEKHVLTAAHCIDG 180
 DB 121 sgksrrkrqiygydsrfsifgkdfllnypfstsvklsctgtvaekhvltaa hcidg 180
 QY 181 KTYVKTOKLRVGFLLKPKKRCGGRANDSTAMPQMKFOWIRVKTVPVPGWIKGNAND 240
 DB 181 ktyvktqrlrvgflfkpkktdggrandsampsqmkfgyrlrvkthvpgwtkgnand 240
 QY 241 IGMDDYALIELLKKPKRRKFMKIGVSPPAKQIPGRIHFSGYDNDRCNLYRRCVYKDE 300
 DB 241 igmdyalyalllkkpkrrkfmki gvspakqipggrihfsydnndrcnlyr rcdvkd 300
 QY 301 TYDLTYOCCDAOPGASGSGVYVYRMKROQKWERKIIIGFSGHQWVDMNGSPDDFNVAVR 360
 DB 301 tydllyqccdaqpgasgsqgyvyrvmkrtqqkwerkiiigifsgbqwdmngspddfnvavr 360
 QY 361 ITPLKYAQICWIKGNVLDCREG 383
 DB 361 itplkyaqicwyikgnvldcreg 383

RESULT 3

ID AAY13390 standard; Protein; 383 AA.

XX AAV13390;
 AC 25-JUN-1999 (first entry)
 DT
 XX Amino acid sequence of protein PRO307.
 DE
 XX Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibrinogen; dermal scarring; Usher Syndrome; Atrophica areata;
 KW anti-thrombotic; wound healing; tissue repair.
 OS Homo sapiens.
 XX MO9914328-A2.
 XX 25-MAR-1999.
 XX 16-SEP-1998; 98WO-US19330.
 PE 25-NOV-1997; 97US-0066840.
 XX 17-SEP-1997; 97US-0059113.
 PR 17-SEP-1997; 97US-0059115.
 PR 17-SEP-1997; 97US-0059117.
 PR 17-SEP-1997; 97US-0059119.
 PR 17-SEP-1997; 97US-0059121.
 PR 17-SEP-1997; 97US-0059122.
 PR 17-SEP-1997; 97US-0059184.
 PR 18-SEP-1997; 97US-0059263.
 PR 18-SEP-1997; 97US-0059266.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 21-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0062816.
 PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 28-OCT-1997; 97US-0063564.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.

XX (GETH) GENENTECH INC.
 PA
 XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 PI WPI: 1999-229533/19.
 XX DR N-PSDB; AAX52261.
 XX New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 PS Claim 12, Fig 96; 320pp; English.
 XX AAV13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophica areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX
 SQ Sequence 383 AA:
 Query Match 100.0%; Score 2080; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 8e-147;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGIGLLEFLLCAVGVSPYSAPMKPTWPAVRLPVLPQSTLNIAKDPFGAEAKLE 60
 DB 1 magipgllfllfllcavgvspysapwkpwpayrlpvlpqstlniakdpfgaeakle 60
 QY 61 VSSSCGPOCHGTRPPTPEBAKQVLSYETLVANGSRITFOVCVITLSSGCGAORHDSGS 120
 DB 61 vssscgpcqchkgtrpptyeekqylsyetlvangsrirtetqvqyillsdgdqqrhds 120
 QY 61 VSSSCGPOCHGTRPPTPEBAKQVLSYETLVANGSRITFOVCVITLSSGCGAORHDSGS 120
 DB 61 vssscgpcqchkgtrpptyeekqylsyetlvangsrirtetqvqyillsdgdqqrhds 120
 QY 121 SGKSRKROIVGYDSRFSGIFGKDFLLNPFSTSVTLSTGCTGTVAEKHVLTAACIHDS 180
 DB 121 sgkstrkrqlygdsrfsifgkdfllnpyfstrsvtlstgctgtvaekhvltacihds 180
 QY 181 KTYVGTGKLRVGLTKPKFKDGRGANDSTSAPEQMKFQWIRKRVHPKGIKGNAND 240
 DB 181 ktyvgtgklrvgltkpkfkdgrgandsstaepemkfqwlrvkrthvpxgwlkgnand 240
 QY 241 IGMDYVALLELKKRHKRRPKMKIGVSPAKOLPGGRHIFSGDNDPRLNLYRPFCDVDE 300
 DB 241 igmdyvallelkkpkrkrrfmklyvspakqlpggrhlifsgdndrplnlyrpfcdvde 300
 QY 301 TYDLLYQOCDAOPGASGSGVYVRMKRROQKWKRIIGIFSGHGVNDKNGSPDFNVAVR 360
 DB 301 tydlllyqgcdaopgassggyvrmkrrqgqkwertllifsgbhgvndnngspdfnva 360
 QY 361 ITPLKYAQCITWIKGNVYLDCEG 383
 DB 361 itplkyaqicwyikgnvlyldcreg 383
 RESULT 4
 AAB25592

ID AAB25592;standard; Protein; 383 AA.
 XX
 AC AAB25592;
 DE
 DT 21-NOV-2000 (first entry)
 XX
 Protein encoded by human secreted protein gene #10 clone HUS005.
 XX
 OS Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 XX antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
 XX anticancer; vulnery; antiviral; antibacterial; antifungal;
 XX Immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 XX Crohn's disease; nephritis; hyperproliferative disorder;
 XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 XX melanoma; lymphoma; wound healing; human; chromosome 12.
 XX
 XX Homo sapiens.
 OS
 XX MO200029435-A1.
 PN
 XX 25-MAY-2000.
 PD
 XX 27-OCT-1999; 99MO-US25031.
 XX
 PR 28-OCT-1998; 98US-0105971.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI M J, Ruben SM, Olsen HS, Young PE, Kenny JT, Moore PA, Wei Y;
 PI Greene JM;
 DR
 XX WPI; 2000-387742/33.
 PT Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases -
 XX
 PS Claim 1; Page 684-685; 803pp; English.
 XX
 The present invention relates to 12 secreted human proteins and the
 nucleotide sequences encoding them. The polynucleotide sequences given
 in AAB80606-A80623 encode the 12 secreted protein sequences given in
 AAB25576-B25593. The human secreted proteins have various activities
 dependent on the tissues in which they are expressed. Examples of the
 activities of the proteins include: immunosuppressant;
 anti-inflammatory; antiarthritic; antirheumatic; dermatological;
 antiproliferative; antiarteriosclerotic; anticancer; vulnery;
 antiviral; antibacterial; and antifungal activity. The proteins,
 polypeptides, agonists and antagonists may be used to treat prevent
 and/or diagnose various disease, disorders and conditions examples of
 which include: immune disorders e.g. Addison's disease, rheumatoid
 arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 hyperproliferative disorders such as paraneoplasias and purpura;
 cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 sequences may also be used in wound healing and the treatment of
 infectious diseases. The human secreted protein gene #10 and protein
 sequences are represented in sequences AAB80615 and AAB25585. Secreted
 protein gene #10 is located on chromosome 12. Sequences AAB25616-B25618
 represent alternative secreted protein #10 sequences and AAB80669-A80676
 represent genes which are related to the secreted protein gene#10.
 XX
 Sequence 383 AA;
 XX

[illegible]

in AAA80606-A80623 encode the 12 secreted protein sequences given in AAB25376-B25393. The human secreted proteins have various activities dependent on the tissues in which they are expressed. Examples of the activities of the proteins include: immunosuppressant; anti-inflammatory; antirheumatic; antineoplastic; dermatological; antiproliferative; antileukemic; anticancer; antitumor; antiviral; antibacterial; and antifungal activity. The proteins, polypeptides, agonists and antagonists may be used to treat prevent and/or diagnose various diseases, disorders and conditions examples of which include: immune disorders e.g. Addison's disease, rheumatoid arthritis, dermatitis, and multiple sclerosis; inflammatory disorders e.g. inflammatory bowel disease, Crohn's disease and nephritis; hyperproliferative disorders such as paraneoplasia and purpura; cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The proteins and polynucleotide sequences may also be used in wound healing and the treatment of infectious diseases. The human secreted protein gene #10 and protein sequences are represented in sequences AAA80615 and AAB25585. Secreted protein gene #10 is located on chromosome 12. Sequences AAB25618-B25618 represent alternative secreted protein #10 sequences and AAA80669-A80676 represent genes which are related to the secreted protein gene#10.

Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 21; Length 383;
Best Local Similarity 100.0%; Pred. No. 8e-147;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAGIGLFLFLFLLCAVGOVSPYAPKPTPARLRYVLPQSTLNIAKPPFGAEAKLE 60
1 magigp11fllfllcavqvpsypapkwptparlrylvlpqstlniakppdfigaeakle 60
61 VSSSCGPOCHKGTPLPYTEAKOYLSEETLYANGSRTETOVGITYLSSSGDAQRDSCS 120
61 vssscgpcqchkgtplyteeakqylsyetlyangsertetovqiyillysssgdaqrdscs 120
121 SSKSRKROIGYDSRFSIFGKDFLLNPFSTSVKLSFGCTGLVAEKHVLTAACHIDG 180
121 sgksrrkroigydsrfsifgkdfllnfpstsvklsfgctglvaekhvltachidg 180
131 KTYVGTOKLRVGFELKPKFGKGRANDSTAMPOMKFRVTRVTHVPRKMTIGNAND 240
131 ktyvgtklrvlgfllkpkfkdggrandstampomkfrvtrvthvprkxwlgignand 240
141 KTYVGTQGLVIRVGLIKPKFKDGGRANDSTAMPOMKFRVTRVTHVPRKMTIGNAND 240
141 ktyvgtgqlrvlgllkpkfkdggrandstampomkfrvtrvthvprkxwlgignand 240
241 IGMDDYVALLLELKKPKRRKFMKIGVSPPAKOLPGRIHFSGYDNDPGLNLYRRCVDE 300
241 igmddyalldelkkpkrrkfmkigvspapakolpgrihfsgydndprpnllyrrcvde 300
301 TYDLYOCCDAOPGASGSGVYVMMKROQKWERKIIIGIFSGHWNVMNGSPDEFNAVR 360
301 tydlyoccdapgasgsgvyvmmkroqkwerkiigifsghwnvmngspdefnavr 360
361 ITPLYAOICVYIKGNVLDRCRC 383
361 itplyaoicvyikgnvldrcrc 383

RESULT 6

AA88277 standard; Protein: 383 AA.

AA88277;

16-OCT-2000 (first entry)

Human TANGO 186 protein.

TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human; murine;
secreted protein; transmembrane protein; gene therapy; vaccine;
diagnosis; treatment; detection.

OS Homo sapiens.
XX WO200018904-A2.
XX 06-APR-2000.
XX 30-SEP-1999; 99WO-US22817.
XX 30-SEP-1998; 98US-0164220.
XX 02-OCT-1998; 98US-0164169.
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX Barnes TM;
XX WPI: 2000-293144/25.
XX N-PSDB: AAA39945; AAA39946.
XX Isolated nucleic acids encoding TANGO polypeptides useful for
XX preventing, diagnosing and treating diseases associated with
XX inappropriate protein expression -
XX
XX
XX Claim 9; Fig 13; 249pp; English.

This invention describes novel human and murine nucleic acids encoding TANGO polypeptides (which are either wholly secreted or transmembrane proteins) which can be used for gene therapy and/or vaccination. The peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215 polypeptides according to standard recombinant DNA methodologies. They may also be used to detect and quantify the presence of TANGO nucleic acids in a sample and therefore identify or diagnose diseases associated with inappropriate TANGO expression (e.g. diseases related to over or under expression of the polypeptides or the expression of inactive polypeptides). The nucleic acids and the polypeptides they encode may be used according to standard gene therapy protocols, to treat diseases associated with inappropriate TANGO expression by supplementing a patient's own production of an abnormally active polypeptide. The may result in expression of an abnormally active polypeptide. The polypeptides may also be used to identify and produce agonists and antagonists of TANGO expression and activity which may be used to modulate TANGO related processes and diseases. The polypeptides are particularly useful for use as antigens for producing antibodies to TANGO proteins which may be used for inhibiting the activity of TANGO proteins. They may also be used to detect and quantify the presence of TANGO proteins in samples and therefore identify patients in whom the protein is over- or under-expressed. This sequence represents the human TANGO 186 protein described in the method of the invention.

Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 21; Length 383;
Best Local Similarity 100.0%; Pred. No. 8e-147;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAGIGLFLFLFLLCAVGOVSPYAPKPTPARLRYVLPQSTLNIAKPPFGAEAKLE 60
1 magigp11fllfllcavqvpsypapkwptparlrylvlpqstlniakppdfigaeakle 60
61 VSSSCGPOCHKGTPLPYTEAKOYLSEETLYANGSRTETOVGITYLSSSGDAQRDSCS 120
61 vssscgpcqchkgtplyteeakqylsyetlyangsertetovqiyillysssgdaqrdscs 120
121 SSKSRKROIGYDSRFSIFGKDFLLNPFSTSVKLSFGCTGLVAEKHVLTAACHIDG 180
121 sgksrrkroigydsrfsifgkdfllnfpstsvklsfgctglvaekhvltachidg 180
131 KTYVGTOKLRVGFELKPKFGKGRANDSTAMPOMKFRVTRVTHVPRKMTIGNAND 240
131 ktyvgtklrvlgfllkpkfkdggrandstampomkfrvtrvthvprkxwlgignand 240
141 KTYVGTQGLVIRVGLIKPKFKDGGRANDSTAMPOMKFRVTRVTHVPRKMTIGNAND 240
141 ktyvgtgqlrvlgllkpkfkdggrandstampomkfrvtrvthvprkxwlgignand 240
241 IGMDDYVALLLELKKPKRRKFMKIGVSPPAKOLPGRIHFSGYDNDPGLNLYRRCVDE 300

useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung squamous cell carcinoma), gastrointestinal disorders (e.g. enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease), wound repair, cardiovascular disorders (e.g. endometrial bleeding angiogenesis, ischaemia such as coronary ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma, rheumatoid arthritis, multiple sclerosis), infertility, AIDS and diabetes and retinal disorders such as retinitis pigmentosa. The PRO nucleic acids have applications in molecular biology, including use as hybridization probes, and in chromosome and gene mapping.

Sequence 383 AA:

Query Match 100.0%; Score 2080; DB 22; Length 383;
Best Local Similarity 100.0%; Pred. No. 8e-147;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLLCAVGVSPYSAPFWKPTWPAVRLPVVLPQSTINLAKPDGAERKLE 60
DB 1 magipglflflflfllcavgvspysapwkpwpayrlpvvlpqstlnlakpdigaekle 60
61 VSSSCGQCHKGMPLEPYEEAKOYLYETLYANGSRTEOVGIYILSSGDAQHRDGS 120
DB 61 vssscgpcqchkgpripyeekqylsyetlyangsrteovgiyilssgdqghrds 120
121 SGKSRKROLYGYDSRFSIFGKDFLNYPESTSVKLSGCTGTVAEKHVLTAACIHG 180
DB 121 sgksrrkrolygydsrfsifgkdfllnypstsvklsctgtvlaekhvltaaahchdg 180
QY 181 KTYVKGQKLAVGFLKRFKDGGRGANDSTSAMPEQMKFQIRKTHVKGIRKGNAND 240
DB 181 ktyvkgqklavgflkrfkdggrgandstsampeqmkfqlvkrthvpgyikgnand 240
QY 241 IGMVDYALLELKKPHKRFKMGKIGVSPPAKQLPGRIHFSGYDNDPGLVYRFDVKE 300
DB 241 igmvydylallelkkphkrfkmgkigvspakqlpggrihfsgydnrdpnlvyrfcdvke 300
QY 301 TYDLLYQCDAPGASGSgyvYRMKRRQOKWERKIIGIFSGHGWDMNGSPDENVAVR 360
DB 301 tydillyqcdapgasgsygyvYRMKRRQOKWERKIIGIFSGHGWDMNGSPQDINAVR 360
QY 361 ITPKYAQCWIKGNLYDCREG 383
DB 361 itplkyaqicwlykgnlydcreg 383

RESULT 10

AAB48974
ID AAB48974 standard; Protein: 383 AA.

AAB48974:

27-MAR-2001 (first entry)

Human zsig13 variant #3, SEQ ID NO:18.

Human zsig13: serine protease; chromosome 11q22.1; elastase homologue;
glutaryl endopeptidase homologue; factor X homologue; trypsin homologue;
trypsin homologue; mast cell protease homologue;
collagenase homologue; protein degradation; food processing; brewing;
alcohol production; laundry detergent component.

Homo sapiens.

US6153420-A.

28-NOV-2000.

04-MAY-1998; 98US-0072384.

24-APR-1997; 97US-0044185.
17-APR-1998; 98US-0062142.

XX (ZYWO) ZYMOGENETICS INC.

PI Sheppard PO;

DR WPI: 2001-060090/07.

XX N-PDB: AAC91784.

PT New isolated serine protease (designated zsig13), useful in industrial
PT processes to degrade unwanted proteins or alter the characteristics of
PT protein-containing composition, as well as in industrial applications
PT (e.g. brewing)

PS Claim 1; Column 41-44; 26pp; English.

CC The invention relates to human zsig13 proteins (AAB4897-BA8974), and
CC to DNA encoding them (AAC91782-C91784). The invention also relates to
CC expression vectors and host cells comprising a human zsig13 DNA, and the
CC recombinant production of a human zsig13 protein or its precursor.
CC zsig13 is a serine protease, and has significant homology to Bacillus
CC licheniformis glutaryl endopeptidase, human clotting factor X, human
CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine
CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human
CC zsig13 is located on chromosome 11q22.1. zsig13 is useful in industrial
CC processes to degrade unwanted proteins or alter the characteristics of
CC protein-containing compositions. It may also be used in industrial
CC applications in which proteases are utilised, including food processing,
CC brewing and alcohol production, and as a component of a laundry
CC detergent. The present sequence represents a human zsig13 variant.

Sequence 383 AA:

Query Match 100.0%; Score 2080; DB 22; Length 383;
Best Local Similarity 100.0%; Pred. No. 8e-147;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLLCAVGVSPYSAPFWKPTWPAVRLPVVLPQSTINLAKPDGAERKLE 60
DB 1 magipglflflflfllcavgvspysapwkpwpayrlpvvlpqstlnlakpdigaekle 60
61 VSSSCGQCHKGMPLEPYEEAKOYLYETLYANGSRTEOVGIYILSSGDAQHRDGS 120
DB 61 vssscgpcqchkgpripyeekqylsyetlyangsrteovgiyilssgdqghrds 120
121 SGKSRKROLYGYDSRFSIFGKDFLNYPESTSVKLSGCTGTVAEKHVLTAACIHG 180
DB 121 sgksrrkrolygydsrfsifgkdfllnypstsvklsctgtvlaekhvltaaahchdg 180
QY 181 KTYVKGQKLAVGFLKRFKDGGRGANDSTSAMPEQMKFQIRKTHVKGIRKGNAND 240
DB 181 ktyvkgqklavgflkrfkdggrgandstsampeqmkfqlvkrthvpgyikgnand 240
QY 241 IGMVDYALLELKKPHKRFKMGKIGVSPPAKQLPGRIHFSGYDNDPGLVYRFDVKE 300
DB 241 igmvydylallelkkphkrfkmgkigvspakqlpggrihfsgydnrdpnlvyrfcdvke 300
QY 301 TYDLLYQCDAPGASGSgyvYRMKRRQOKWERKIIGIFSGHGWDMNGSPDENVAVR 360
DB 301 tydillyqcdapgasgsygyvYRMKRRQOKWERKIIGIFSGHGWDMNGSPQDINAVR 360
QY 361 ITPKYAQCWIKGNLYDCREG 383
DB 361 itplkyaqicwlykgnlydcreg 383

RESULT 11

AAB48973
ID AAB48973 standard; Protein: 392 AA.

AAB48973:

27-MAR-2001 (first entry)

XX Human Zs1g13 variant #2, SEQ ID NO:15.
 DE
 XX
 KM Human Zs1g13; serine protease; chromosome 11q22.1; elastase homologue;
 KM glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;
 KM trypsinogen homologue; mast cell protease homologue;
 KM collagenase homologue; protein degradation; food processing; brewing;
 KM alcohol production; laundry detergent component.
 OS Homo sapiens.
 PN US6153420-A.
 XX
 XX 28-NOV-2000.
 PD
 XX
 PF 04-MAY-1998; 98US-0072384.
 XX
 XX 24-APR-1997; 97US-0044185.
 XX 17-APR-1998; 98US-0062142.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Sheppard PO;
 PI
 DR WPI: 2001-060090/07.
 DR N-PSDB: AAC91783.
 XX
 PT New isolated serine protease (designated Zs1g13), useful in industrial
 PT processes to degrade unwanted proteins or alter the characteristics of
 PT protein-containing composition, as well as in industrial applications
 PT (e.g. brewing).
 PS
 PS Claim 1; Column 35-38; 26pp; English.
 XX
 CC The invention relates to human Zs1g13 proteins (AAB48972-B48974), and
 CC to DNA encoding them (AAC91782-C91784). The invention also relates to
 CC expression vectors and host cells comprising a human Zs1g13 DNA, and the
 CC recombinant production of a human Zs1g13 protein or its precursor.
 CC Zs1g13 is a serine protease, and has significant homology to Bacillus
 CC licheniformis glutamyl endopeptidase, human clotting factor X, human
 CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine
 CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human
 CC Zs1g13 is located on chromosome 11q22.1. Zs1g13 is useful in industrial
 CC processes to degrade unwanted proteins or alter the characteristics of
 CC protein-containing compositions. It may also be used in industrial
 CC applications in which proteases are utilized, including food processing,
 CC brewing and alcohol production, and as a component of a laundry
 CC detergent. The present sequence represents a human Zs1g13 variant.
 XX
 SQ Sequence 392 AA:
 Query Match 100.0%; Score 2080; DB 22; Length 392;
 Best Local Similarity 100.0%; Pred. No. 8.3e-147;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 241 IgmDyDyAlleIkphkRrkImkiyvspakqIpggrIhfsyDndrpnIlyrfcdvxc 300
 QY 301 FYDILYQOCDAOPGASGSgyVRMkRROOKWERKITGFSCHOWNMNGSPQDENVAVR 360
 DB 301 tYdIllyqcdagpagsgyvymkrrqgkwerkiIgfsgywmngspqdnvaVR 360
 QY 361 tPELRYAOICWIKGNyLDCREG 383
 DB 361 tPELRYAOICWIKGNyLDCREG 383
 RESULT 12
 ID AAB88336 standard; Protein: 383 AA.
 XX
 XX AAB88336;
 AC
 XX 23-MAY-2001 (first entry)
 DT
 XX
 DE Human membrane or secretory protein clone PSEC0048.
 XX
 KM Human; secretory protein; membrane protein; vaccine; gene therapy;
 KM rheumatoid arthritis; diabetes.
 OS Homo sapiens.
 PN EPI067182-A2.
 XX
 PD 10-JAN-2001.
 PF
 PF 07-JUL-2000; 2000EP-0114090.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 DR WPI: 2001-093989/11.
 DR N-PSDB: AAF93763.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 PS
 PS Claim 1; SEQ ID 40; 609pp + CD ROM; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.

Query Match	Best Local Similarity	Score	Length	DB	21:	Score	Length	DB	21:
Matches 354; Conservative	5;	Mismatches	7;	Indels	1;	Gaps	14;		
1	MAGIGGLFLFLFFLLCAVGOVSPYAPAMPKPTPAVRLPVYLPOSTLNLAKPFGAEAKLE	60							
1	magigglflflflflcavqavspysapwlpvayrlpvlpdstlnlakpfdgaekle	60							
61	VSSSCGPCCHKGTPLPTVEEAKOYLSYERLVANGSRRTETOVGIYLLSSGGDAORHDSGS	120							
61	vssscgpcchkgtpptveeakqyslyellvngsrtrtetqvgyillylssggdgaqnrhdsgs	120							
121	SGSKRRKROIVGYDSRFSIFGKDFLLNTPFSTSVKLISTGCTGTLVAEKHVLTAHCHIDG	180							
121	sgskrrkrqivgydsrfsifgkdfllnypfstsvklistgctgtlvaekhvltaachidg	180							
181	KTYVGTQKLRNGFLKPKRKDGGRGANSTAMPQMKQWIRVTRTHYPKWIGNAD	240							
181	ktyvgtgqlrvngflkpkrkdggrgandstampeqmkqwtvtrvtrthbpxkwlignad	240							

QY		IGMDVALLLEKRRKHKKRPFMKIGSPAPAKOLPGSRHIFSGYDNDRGNNLYRFCDVKDE	300
Dd		241 IGMDDVALLEKRRKHKKRPFMKIGSPAPAKLPGRIRHIFSGYDNDRGNNLYRFCDVKDE	300
QY		301 TYDLTYQQCDAOPGASGSgyvVRMWRKROOQRWERRIKGIFSGHOMVDNMGSPDPFNNAV-	359
Dd		301 tydlItyqgcdaopgassgsyvyvrwmkrqhgwerriklglmslsghwvdmdsgspgfufgcs	360
QY		360 RTPLPKY 366	
Dd		361 eItPlqY 367	
RESUFT_15			
AAAB25617	ID	AAAB25617 standard; Protein: 413 AA.	
AC	AAAB25617;		
DT	21-NOV-2000	(first entry)	
XX		Protein encoded by human secreted protein gene #10.	
DE		Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;	
KW		anti rheumatic; dermatological; antiproliferative; antiarteriosclerotic;	
KW		anticancer; vulnery; antiviral; antibacterial; antifungal;	
KW		immune disorder; Addison's disease; Rheumatoid arthritis; dermatitis;	
KW		Crohn's disease; nephritis; hyperproliferative disorder;	
KW		cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;	
KM		/ melanoma; lymphoma; wound healing; human; chromosome 12.	
OS		Homo sapiens.	
XX		WO200029435-A1.	
PN		25-MAY-2000.	
PD		27-OCT-1999; 99WO-US25031.	
PE		28-OCT-1998; 98US-0105971.	
PR		(HUMA-) HUMAN GENOME SCI INC.	
PA		Ni J., Ruben SM., Olsen HS., Young PE., Kenny JJ., Moore PA., Wei Y.;	
PI		Greene JM;	
DR		WPI: 2000-387742/33.	
XX		The present invention relates to 12 secreted human proteins and the	
CC		nucleotide sequences encoding them. The polynucleotide sequences given	
CC		in AAAA06066-880623 encode the 12 secreted protein sequences active in	
CC		AAA25576-B25593. The human secreted proteins have various activities	
CC		dependent on the tissues in which they are expressed. Examples of the	
CC		activities of the proteins include: immunosuppressant;	
CC		anti-inflammatory; antiarthritic; anti rheumatic; dermatological;	
CC		antiproliferative; antiarteriosclerotic; anticancer; vulnery;	
CC		antiviral; antibacterial; and antifungal activity. The proteins,	
CC		polypeptides, agonists and antagonists may be used to treat prevent	
CC		and/or diagnose various disease, disorders and conditions examples of	
CC		which include: immune disorders e.g. Addison's disease, Rheumatoid	
CC		arthritis, dermatitis, and multiple sclerosis; inflammatory disorders	
CC		e.g. inflammatory bowel disease, Crohn's disease and nephritis;	
CC		hyperproliferative disorders such as paraproteinemias and purpura;	
CC		cardiovascular disorders e.g. coronary arteriosclerosis and myocardiitis	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:39:10 ; Search time 12.21 Seconds
(without alignments)
631.891 Million cell updates/sec

Title: US-09-554-933-3
Perfect score: 2080
Sequence: 1 MAGIPULFLFLCAVQG.....LKYAQCWIKWYLDQREG 383

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4	US-09-072-384-18
2	2080	100.0	392	4	US-09-072-384-15
3	2044	98.3	392	4	US-09-072-384-2
4	177	8.3	222	1	US-08-090-048-1
5	177	8.5	222	2	US-08-292-550-1
6	177	8.5	222	2	US-07-927-661A-1
7	122	5.9	256	3	US-08-906-769-89
8	122	5.9	256	4	US-08-906-616-89
9	122	5.9	256	4	US-08-817-795-89
10	122	5.9	256	4	US-08-639-075A-89
11	122	5.9	256	4	US-09-012-431-89
12	122	5.9	256	4	US-09-032-215-32
13	122	5.9	256	4	US-09-012-692-89
14	122	5.9	256	4	US-08-906-613-89
15	122	5.9	256	4	PT-US95-14442A-89
16	117	5.6	437	1	US-08-487-037-2
17	115.5	5.5	241	4	US-08-944-483-59
18	114	5.5	228	4	US-08-944-483-44
19	114	5.5	253	6	5223425-8
20	113	5.4	238	6	5223425-5
21	113	5.4	250	6	5223425-4
22	112.5	5.4	223	1	US-08-278-091-13
23	112.5	5.4	223	1	US-08-483-859-13
24	112.5	5.4	223	1	US-08-472-173-13
25	112.5	5.4	223	2	US-08-487-167-13
26	112.5	5.4	223	2	US-08-482-816-13
27	112.5	5.4	223	2	US-08-296-149-13

28	112.5	5.4	223	2	US-08-801-499-13	Sequence 13, Appl
29	112.5	5.4	223	2	US-08-615-271-13	Sequence 13, Appl
30	112.5	5.4	223	3	US-09-074-660-13	Sequence 13, Appl
31	112.5	5.4	223	3	US-09-074-659-13	Sequence 13, Appl
32	112.5	5.4	223	4	US-09-106-468-13	Sequence 13, Appl
33	112.5	5.4	223	4	US-09-106-466A-13	Sequence 13, Appl
34	112.5	5.4	223	4	US-09-106-467-13	Sequence 13, Appl
35	111.5	5.4	241	1	US-08-330-978-4	Sequence 4, Appl
36	111.5	5.4	241	1	US-08-474-042-4	Sequence 4, Appl
37	111.5	5.4	241	1	US-08-484-558-4	Sequence 4, Appl
38	111.5	5.4	241	1	US-08-774-592-4	Sequence 4, Appl
39	111.5	5.4	254	1	US-08-330-978-3	Sequence 3, Appl
40	111.5	5.4	254	1	US-08-474-042-3	Sequence 3, Appl
41	111.5	5.4	254	1	US-08-484-558-3	Sequence 3, Appl
42	111.5	5.4	254	1	US-08-774-592-3	Sequence 3, Appl
43	111.5	5.4	306	1	US-08-330-978-1	Sequence 1, Appl
44	111.5	5.4	306	1	US-08-474-042-1	Sequence 1, Appl
45	111.5	5.4	306	1	US-08-484-558-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-072-384-18
Sequence 18, Application US/09072384
Patent No. 6153420
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072.384
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E 31,648
REGISTRATION NUMBER: 97-16C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1-19
OTHER INFORMATION:
US-09-072-384-18

```

Query Match          100.0%; Score 2080; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 4,6e-223;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MAGIDPELLFLLPILCAVGQVSPYSAPMKPTWPAVRLPVYLPOSTLNLAKDPFGAEAKLE 60
DB      1  MAGIDPELLFLLPILCAVGQVSPYSAPMKPTWPAVRLPVYLPOSTLNLAKDPFGAEAKLE 60
QY      61  VSSSGPCCHGHTPLPTVEEAKQVLSYETLVANGSRTETQVGIYLLSSSGDAQHRDSSG 120
DB      61  VSSSGPCCHGHTPLPTVEEAKQVLSYETLVANGSRTETQVGIYLLSSSGDAQHRDSSG 120
QY      121  SGKSRKRRQIYGYDSRFSIFGKDFLLNLPSTSVKLSFGCTGLVAEKHVLTAHCIDHG 180
DB      121  SGKSRKRRQIYGYDSRFSIFGKDFLLNLPSTSVKLSFGCTGLVAEKHVLTAHCIDHG 180
QY      181  KTYVKGTOKLGVFLKPKFKDGGKANDSTAMPROMQWIRVVRTHPKKWMIGNAND 240
DB      181  KTYVKGTOKLGVFLKPKFKDGGKANDSTAMPROMQWIRVVRTHPKKWMIGNAND 240
QY      241  IGMIDYVALLLEKKPKHKKRPMKIGVSPPAKOLPGRIHFSGYDNDPQNLVYRCDVKE 300
DB      241  IGMIDYVALLLEKKPKHKKRPMKIGVSPPAKOLPGRIHFSGYDNDPQNLVYRCDVKE 300
QY      301  TYDILYQOCDAOPGSSGGVYVRMKRQOKKERKIIIGFSGHOWYDMNGSPQDFNVAVR 360
DB      301  TYDILYQOCDAOPGSSGGVYVRMKRQOKKERKIIIGFSGHOWYDMNGSPQDFNVAVR 360
QY      361  ITPLEKAOICYWKIKGNVILDCREG 383
DB      361  ITPLEKAOICYWKIKGNVILDCREG 383

RESULT 2
US-09-072-384-15
: Sequence 15, Application US/09072384
: Patent No. 6153420
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES
: TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ZymoGenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/072,384
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, Gary E
: REGISTRATION NUMBER: 31,648
: REFERENCE/DOCKET NUMBER: 97-16C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-442-6673
: TELEFAX: 206-442-6678
: TELEX:
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 392 amino acids

```

```

: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: FEATURE:
: NAME/KEY: Signal Sequence
: LOCATION: 1...19
: OTHER INFORMATION:
:
: US-09-072-384-15
:
Query Match      100.0%; Score 2080; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 4.8e-223;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
:
QY 1 MAGTIGLFLFELLCAVGQSPYSAPMKPTWPAVRLPYVLPOSTLNLAKPFGAEAKLE 60
Db 1 MAGTIGLFLFELLCAVGQSPYSAPMKPTWPAVRLPYVLPOSTLNLAKPFGAEAKLE 60
:
QY 61 VSSSGPCQCHKGTPLPTYEAKQVLSYETLVANGSRTETQVGIYLLSSGDAQHRDSS 120
Db 61 VSSSGPCQCHKGTPLPTYEAKQVLSYETLVANGSRTETQVGIYLLSSGDAQHRDSS 120
:
QY 121 SKSRKRRKQIYGYDSRFSTFGKDFLLNTPFSTSVKLTSCCTGTLVAEKVLLAHCIHG 180
Db 121 SKSRKRRKQIYGYDSRFSTFGKDFLLNTPFSTSVKLTSCCTGTLVAEKVLLAHCIHG 180
:
QY 181 KTYVGTQKLRVGFLEPKFKDGGRGANDSTSMPEQMKQWIRKTHVPKGMKIGNAND 240
Db 181 KTYVGTQKLRVGFLEPKFKDGGRGANDSTSMPEQMKQWIRKTHVPKGMKIGNAND 240
:
QY 241 IGMDDYVALLLEKTKHKKRFFMKIGVSPRAKQLPGGIHFSGYDNDPGLVTRFCQYKDE 300
Db 241 IGMDDYVALLLEKTKHKKRFFMKIGVSPRAKQLPGGIHFSGYDNDPGLVTRFCQYKDE 300
:
QY 301 TYDLLYQCCDAQPGASGSGVYVRMMKROQOKMERKIIIGIFSGHOWDMANGSPQDFNVAVR 360
Db 301 TYDLLYQCCDAQPGASGSGVYVRMMKROQOKMERKIIIGIFSGHOWDMANGSPQDFNVAVR 360
:
QY 361 ITPKYAQICYWKIGNYLDCCREG 383
Db 361 ITPKYAQICYWKIGNYLDCCREG 383
:
RESULT 3
US-09-072-384-2
: Sequence 2, Application US/09072384
: Patent No. 6153420
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES.
: TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Zymogenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0.
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/072,384
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:

```

ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31,648
REFERENCE/DOCKET NUMBER: 97-16C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Signal Sequence
LOCATION: 1...19
OTHER INFORMATION:
US-09-072-384-2

Query Match 98.3%; Score 2044; DB 4; Length 392;

Best Local Similarity 98.4%; Pred. No. 4.9e-219;
Matches 377; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFFLLCAAGVSPYSAPMKPTMPAYRLPVYLPOSTLNLAKPDGAEAKLE 60
DB 1 MAGIPGLFLFFLLCAAGVSPYSAPMKPTMPAYRLPVYLPOSTLNLAKPDGAEAKLE 60

QY 61 VSSCGPCHKGTPLPYEAQYLYETLYANGSRTETQVGIYLLSSGDAQHRDSCS 120
DB 61 VSSCGPCHKGTPLPYEAQYLYETLYANGSRTETQVGIYLLSSGDAQHRDSCS 120

QY 121 SGSKRRKQIYGYSRSEIFPKDFLNPSTSVKLTGCTGLVAEKHVLTAACIHG 180
DB 121 SGSKRRKQIYGYSRSEIFPKDFLNPSTSVKLTGCTGLVAEKHVLTAACIHG 180

QY 181 KTYVGTOKLAVGLPKPKFGKGRANDSTAMPQOMFQWIRVKTVPKGIKGNAND 240
DB 181 KTYVGTOKLAVGLPKPKFGKGRANDSTAMPQOMFQWIRVKTVPKGIKGNAND 240

QY 241 IGMDDYVALLLEKRPKPKFGKGRANDSTAMPQOMFQWIRVKTVPKGIKGNAND 300
DB 241 IGMDDYVALLLEKRPKPKFGKGRANDSTAMPQOMFQWIRVKTVPKGIKGNAND 300

QY 301 TYDLYOQCDAPGASGVYVRRMKRQOKWERKIIIGFSGHGVDMNGSPDFENAVR 360
DB 301 TYDLYOQCDAPGASGVYVRRMKRQOKWERKIIIGFSGHGVDMNGSPDFENAVR 360

QY 361 ITPLKVAQICWIKGNVLDRCRG 383
DB 361 ITPLKVAQICWIKGNVLDRCRG 383

RESULT 4

US-08-090-048-1

Sequence 1, Application US/08090048

Patent No. 5523237

GENERAL INFORMATION:

APPLICANT: Budtz, Peter

APPLICANT: Nielsen, Per M.

TITLE OF INVENTION: PROTEIN PREPARATIONS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5523237 disk of No. 5523237th America, Inc.

STREET: 405 Lexington Avenue, 62nd Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/090,048
FILING DATE: 16-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 199/91
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK92/00036
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Agtis, Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3396, 214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-090-048-1

Query Match 8.5%; Score 177; DB 1; Length 222;

Best Local Similarity 25.5%; Pred. No. 1.6e-11;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 9;

QY 148 YPFSVSLST---GCTTVLAEKHVLTAACIHGDKT-YKGTOKLAVGLPKPKFGK 203
DB 17 YPRAIVHSSISGCTGWMIGPTVATAGHCTVDTSSGFAAGTAVSPG----- 66

QY 204 RGANDSTAMPQOMFQWIRVKTVPKGIKGNANDIGMDYVALLLEKRPKPKFGK 263
DB 67 ---RNGTS-----YPGVSSTRFTIPSGMKSGNTN-----YDGAITLSEPIGNTYIF 113

QY 264 GVSPPAKOLPGRIHFSGYDNDPRGNLVYRCD--VKDETLYOQCDAPGASGVY 321
DB 114 GYSTTSLVGTVYIISGPDKTAGTQWQSGPIAISEYTKLOYAM-DYVGSGSGSPVF 172

QY 322 VRMKRQOKWERKII---GIFSGHGVDMNGSPDFENAVRITPLKVAQICW 372
DB 173 EOSSRTNCSGPCSLAVHTNGVYG-----SSYNRGTRTKEVEFDMITWN 217

RESULT 5

US-08-292-550-1

Sequence 1, Application US/08292550

Patent No. 5863573

GENERAL INFORMATION:

APPLICANT: Dammann, Claus

APPLICANT: Budtz, Peter

TITLE OF INVENTION: A PROCESS FOR PRODUCING CHEESE

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5863573 disk of No. 5863573th America, Inc.

STREET: 405 Lexington Avenue, Suite 6400

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10174-6401

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

TELEPHONE: (303) 863-37
TELEFAX: (303) 863-0223

LEADER:
INFORMATION FOR SEO

SEQUENCE CHARACTERISTICS:

query match	0.00	0.00	0.00
Best Local Similarity	24.18;	Pred. No.	2.8e-05;

	Matches	49;	Conservative	30;	Mismatches	68;	Indels	56;	Gaps	10;
Oy	101	VGIIYLLS--SGDGAQHNDSSGSSGKSRKQIYGYSRFSIPGKDPLNYPSTSVKLST	158							
Db	8	VGLSAVSSYKIKDGLDGRIVGGQADADIAK---YGYQASLQVFNEHF-----	50							
Oy	159	GCTGTLVAEKHVLTAAHCIDHGTIVKGTQKLRVGFLEKPKFKOGGGRANDSTSAMPEQMK	218							
Db	51	CGASILNNYWIWTAHACIYDEFTY----SVRVG---TSFO-GRRGSVHPXAQIIKHHPA	100							
Oy	219	FCWIRVKKRTHVKGWIKGNANDIGMDYDYLELKKP---HKRKFMKIGVSPPAKQLPQG	275							
Db	101	Y-----GWTDI--DMEXALIKRRPRLNNRTVRYVKLIDVGKDMPSG	142							
Oy	276	RIHF-----SGYDNDRPGNLVY	292							
Db	143	ELATVTGWNIGEDDEDEPEQLQY	165							

Search completed: July 12, 2001, 11:43:10
Job time: 240 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:40:15 ; Search time 15.96 seconds
(without alignments)
1827.998 Million cell updates/sec

Title: US-09-554-933-3
Perfect score: 2080
Sequence: 1 MAGIPGLFLFLFLCAVQ.....LKYYQICWIKGNLYDCRGG 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Database: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	194	9.3	316	2 A45134	endopeptidase (EC
2	135.5	6.5	313	2 A35122	metalloprotease
3	120.5	5.8	482	1 EXRT	coagulation factor
4	119	5.7	269	2 A26823	pancreatic elastase
5	117	5.6	522	2 T29767	hypothetical prote
6	115.5	5.6	269	2 B26823	pancreatic elastase
7	115.5	5.6	271	2 A25528	pancreatic elastase
8	115	5.5	258	2 S70439	pancreatic elastase
9	115	5.5	267	4 A56615	probable pancreati
10	114.5	5.5	238	1 TRWV5Y	trypsin-like prote
11	114	5.5	246	1 DBHU	complement factor
12	113	5.4	266	1 ELRT1	pancreatic elastase
13	113	5.4	259	1 TRSMG	pancreatic elastase
14	112.5	5.4	259	1 TRSMG	pancreatic elastase
15	112.5	5.4	273	2 B26823	hypothetical prote
16	112.5	5.4	273	2 B26823	hypothetical prote
17	111.5	5.4	488	1 EXHU	coagulation factor
18	111	5.3	761	2 JC5759	brain-specific ser
19	110	5.3	492	1 EXBO	coagulation factor
20	110	5.3	492	1 EXBO	coagulation factor
21	110	5.3	1582	2 T15308	hypothetical prote
22	109.5	5.3	405	2 T35117	probable secreted
23	108	5.2	583	2 A29154	complement factor
24	108	5.2	786	2 A47547	serine proteinase
25	107.5	5.2	1047	2 A55617	masquerade precurs
26	107.5	5.2	236	2 A28566	T-cell suppressor
27	107.5	5.0	686	1 A59271	Re-activator factor
28	105	5.0	274	2 S40004	trypsin-related pr
29	103.5	5.0	271	1 ELRT2	pancreatic elastase

30	103.5	5.0	416	1 KFB0	coagulation factor
31	102.5	4.9	1238	2 T34929	hypothetical prote
32	100	4.8	272	2 JC4170	trypsin-like prote
33	98.5	4.7	267	2 S40006	trypsin (EC 3.4.21
34	98	4.7	548	2 D82175	probable trypsin v
35	98	4.7	624	2 T02289	probable polygalac
36	97.5	4.7	409	2 T35118	probable secreted
37	97.5	4.7	452	1 A30351	coagulation factor
38	97.5	4.7	747	2 T51579	coagulation factor
39	97	4.7	1019	2 A38738	complement factor
40	96.5	4.6	259	2 S68424	allergen Der f III
41	96.5	4.6	275	2 T46712	factor IX - rabbit
42	96.5	4.6	2145	2 JC4747	adenylate cyclase
43	95.5	4.6	430	1 A24702	serine proteinase
44	95	4.6	782	2 T32155	hypothetical prote
45	94	4.5	268	2 S68825	pancreatic elastase

ALIGNMENTS

RESULT 1
A45134
endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 10-Jun-1993 #sequence.revision 18-Nov-1994 #text.change 15-Oct-1999
C:Accession: A45134; S23078
R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S
J. Biol. Chem. 267, 23782-23788, 1992
A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s
A:Reference number: A45134; MUID:93054737
A:Accession: A45134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KAK>
A:Cross-references: GB:D10060; NID:g216263; PIDN:BA00949.1; PID:d1001415; PID:g21626
A:Experimental source: ATCC 14580
A:Note: Sequence extracted from NCBI backbone (NCBIN:118784, NCBI:118785)
R:Svensden, I.; Bredam, K.
Eur. J. Biochem. 204, 165-171, 1992
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase
A:Reference number: S23078; MUID:92135199
A:Accession: S23078
A:Status: preliminary
A:Molecule type: protein
A:Residues: 95-316 <SVE>
C:Keywords: hydrolase

Query Match 9.3%; Score 194; Length 316;
Best Local Similarity 24.0%; Pred. No. 1.2e-08;
Matches 79; Conservative 43; Mismatches 133; Indels 64; Gaps 15;

QY	63	SSCGPOCKKGPPL--PYEPAKOTLSTETLYANGSKRTQGVGIYIISSSDGAQHRDSC	119
DB	28	AQAAPSPH--TPVSDPSY-KAETSVTPD--NKSDDYGIYSKAFDTGVNENKE	79
QY	120	SSGSRKRRQLYGDSRSIFGKDFLN-----YFESTVYKLSN--GCTGVAAEKV	170
DB	80	KAEEKSPAKAPY--SIKSVIGSDDRTRVNTTAYPRATYIISSSIGSGCTGMIGPKTV	136
QY	171	LTAACHIDGKT-VYGTQKLAVGLKPKFDGSGANDSTSAMPEQMKPQWIRKRTYV	229
DB	137	ATAGCICITDSSGSPAGATVSPG-----RNGTS-----YFVSVASTRFI	178
QY	230	PKGWIKGANDIGMDYDALLLEKPKRKRKMGVSPPAKOLFQGRTHSGYNDPRGN	289
DB	179	PSGMSGNTN-----YDGALEISEPIGNTVGYGYSTYSLSLVGTVTJSGPGDRTAG	233
QY	290	LVYRCD--VKDEYDILLOCCDQAPGASGSGVYVRMKRQOQWKKRII--GIRSGH	343
DB	234	TOMHSGPILAISETYKLOYAM-DTYGGSGSPVEQSSRTNCGSCPLAVHTNGVYG-	291

OY 344 QWYDMNGSPQDENAVAVITPELKYAOICYW 372
 DB 292 -----SSYNKGRITREVEVDNLTNW 311

RESULT 2

A:Accession: A35122
 C:Species: Bacillus subtilis
 C:Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 15-Oct-1999
 C:Accession: A35122; 140010; A69660
 R:Stroma, A.; Rudolph, C.F.; Rado Jr., G.A.; Sullivan, B.J.; Thierault, K.A.; Ally, D.; F.
 J. Bacteriol. 172, 1024-1029, 1990
 A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.
 A:Reference number: A35122; MUID:90130256
 A:Accession: A35122
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-313 <SLD>
 A:Cross-references: GB:L10505; NID:g143209; PIDN:AAA22604.1; PID:g143210; GB:M29036
 R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.
 Gene 70, 351-361, 1988
 A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s

A:Reference number: 139994; MUID:9108019

A:Accession: 140010
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-60, 65, 'L', 67, 'S', 69, 'AQA' <RES>

A:Cross-references: GB:M22916; NID:g143701; PIDN:AAA22832.1; PID:g143702

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallie

lech, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardinois,

A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portella

Rieger, M.; Rivolta, S.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleif, S.; Schrotter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Ser

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, K.; Tosto, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033

A:Accession: A69660

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-313 <KUN>

A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:el182176;

A:Experimental source: strain 168

Genetics:

A:Gene: mpr

C:Keywords: hydrolase

Query Match 6.5%; Score 135.5; DB 2; Length 313;
 Best Local Similarity 22.2%; Pred. NO. 0.00991;
 Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

OY 72 GNPPLPYEAKOYLSELYLWANGSRTEVOGIYLISSSD--GAOHRDSGGSGKRRKQ 129
 DB 29 GVPAAKAENPQTSVNTGKEADTKNQT-----SKADQVSAPVEGTGKTSK----- 75

OY 130 IYGYDSRF-----SIFGKD-----FLLNPEFSVSLST-----GCTGTL 164
 DB 76 LVGGQLELEKNITLOPSSIIIGDERTRLSSTTSFPRATVOLSTKYPTSSYTGCTGL 135

OY 165 VAEKHVLTAAHCII-----DGKTYVKGTOKLKLVGFLKPKFDGGRGAND 208
 DB 136 VNNNTYVTAHCYVDSODHGMASHTTAPGRNSSPYGY----- 175

OY 209 STSAMPEQMKFQWIRVAKRTHVPRKGIKN---ANDIGMDYALALELKKPRKREKIGV 265

DB 176 -SGTFYISVK -GWTESKDTNVDYAIKLSGSPGNTVGH-YCIRRTNSSP-----VGL 225

OY 266 SPFAKQLPGRIHFGSYDNDPRGNLVYRFCDVDEYDLYOQCDAPGASGSYVYRW 325
 DB 226 SSSVYGFPEDKRFGTMWSQDKPIR-----SAETVILFY-TDLYGCGSGSPY----- 272

OY 326 KROQOKWEKRTIGFSGHGWDMNGSPQDENAVAVITPELKYAOICYW 372
 DB 273 -RNSYDTGQFAIAIHT-----NGG-SSYNLGRIVNDVFNNTIOYW 310

RESULT 3

EXPT
 coagulation factor Xa (EC 3.4.21.6) precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Jan-1995 #sequence_revision 07-Feb-1997 #text_change 08-Dec-2000

C:Accession: S49075; J04670; P50191; P50190; 162745

R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.

Thromb. Res. 80, 63-73, 1995

A:Title: Evidence for competition between vitamin K-dependent clotting factors for in

A:Reference number: A58498; MUID:96093366

A:Accession: S49075

A:Molecule type: mRNA

A:Residues: 1-482 <STA1>

A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601

A:Note: submitted to the EMBL Data Library, June 1994

A:Note: neither the complete nucleic acid sequence nor the complete translation are s

Gene 169, 269-273, 1996

A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.

A:Reference number: J04670; MUID:96194815

A:Accession: J04670

A:Molecule type: mRNA

A:Residues: 1-482 <STA2>

A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601

A:Experimental source: Cos-1 cell

R:Enjiyoji, K.; Miyazaki, K.; Kato, H.

J. Biochem. 109, 890-898, 1991

A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat

A:Reference number: P50190; MUID:92041742

A:Accession: P50191

A:Molecule type: protein

A:Residues: 41-56, 'X', 60-65 <ENU1>

A:Accession: P50190

A:Molecule type: protein

A:Residues: 183-186, 'X', 188-207 <ENU2>

A:Reference number: 146196; MUID:94222160

A:Title: Analysis of the partial nucleotide sequences and deduced primary structures

Fur J. Hematol. 52, 162-168, 1994

R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.

A:Accession: 162745

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 295-383, 'G', 385-455 <MUR>

A:Cross-references: GB:D21215; NID:g415309; PIDN:BA04756.1; PID:g455396.

C:Function:

A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the

A:Pathway: blood coagulation

C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol

C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu

F:1-23/Domain: signal sequence #status predicted <PRO>

F:24-40/Domain: propeptide #status predicted <PRO>

F:41-179/Product: coagulation factor X light chain #status predicted <LCH>

F:129-144/Domain: EGF homology <EG1>

F:183-482/Product: coagulation factor X heavy chain #status predicted <HCH>

F:183-221/Domain: activation peptide #status predicted <APY>

F:232-460/Domain: trypsin homology <TRI>

F:46/47, 54, 56, 59, 60, 65, 66, 69, 72, 79/Modified site: gamma-carboxyglutamic acid (Glu) #

F:57-62, 90-101, 95-110, 112-121, 129-140, 136-149, 151-164, 172-340, 238-243, 259-275, 368-402

A;Accession: A00960
A;Molecule type: mRNA

A:Residues: 1-266 <MAC>
 A:Cross-references: GB:V01234; NID:956088; PIDN:CAA24544.1; PID:956089
 R:Largman, C.
 Biochemistry 22, 3763-3770, 1983
 A:Title: Isolation and characterization of rat pancreatic elastase.
 A:Reference number: A20534; MUID:84000385
 A:Accession: A20534

A:Molecule type: protein
 A:Residues: 17-37, 'X', 39-45 <LAR>
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; pancreas; serine proteinase; zymogen
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:17-26/Domain: activation peptide #status predicted <ACT>
 F:27-266/Product: elastase I #status predicted <MPT>
 F:27-259/Domain: trypsin homology <TRY>
 F:71,119,214/Active site: His, Asp, Ser #status predicted

Query Match 5.4%; Score 113; DB 1; Length 266;
 Best Local Similarity 26.7%; Pred. No. 0.056;
 Matches 39; Conservative 23; Mismatches 48; Indels 36; Gaps 7;

128 RQYGVDSRPSFGKDELRYPESTSVKLTG-----CTGTVAEKHVLTAHCIDHDKT 182
 26 RYVGSAFARNNSPQSLQY-----LSGGMYHTCGTLLRRNMVMTAAHCVSQMT 78
 183 YKGTOKLRVGFLEKPKKRGDGSNDSTAMPQMKFQWIRKRVKTHVPGKMGINDIG 242
 79 F-----RVVVG-----DHNLSQNDGTE-----QYVSQKIVHPYV-----NSNNVA 115
 QY 243 MDYVALLL-KKPKRKKPKKIGSP 267
 DB 116 AGYDALLRLAQSVTLNNVQLAVLP 141

RESULT 13

ELRG
 pancreatic elastase (EC 3.4.21.36) I precursor - pig

C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1990 #text_change 16-Jun-2000
 C:Accession: J50013; A26777; A10061; A00959
 R:Shirasu, Y.; Yoshida, H.; Miyayama, T.; Matsuki, S.; Tanaka, J.I.; Ikenaga, H.
 J. Biochem. 99, 1707-1712, 1986
 A:Title: Isolation and expression in *Escherichia coli* of a cDNA clone encoding porcine B
 A:Reference number: A92005; MUID:86304235
 A:Accession: J50013

A:Molecule type: mRNA
 A:Residues: 1-266 <SHI>
 A:Cross-references: GB:X04036; GB:D00070; GB:N00070; NID:91941; PIDN:CAA27670.1; PID:919
 R:Tani, T.; Kawashima, I.; Furukawa, H.; Ohmine, T.; Takiguchi, Y.
 Biochem. 101, 591-599, 1987
 A:Title: Characterization of a silent gene for human pancreatic elastase I: structure of
 A:Reference number: A26777; MUID:87250343

A:Accession: A26777
 A:Molecule type: mRNA
 A:Residues: 1-125, 'G', 127-183, 'L', 185-266 <TAN>
 A:Cross-references: GB:D00160; NID:9217683; PIDN:BA00118.1; PID:9217684
 R:Tani, T.; Kawashima, I.; Furukawa, H.; Ohmine, T.; Takiguchi, Y.
 Biochem. J. 131, 643-675, 1973
 A:Title: Evidence for the amino acid sequence of porcine pancreatic elastase.
 A:Reference number: A90267; MUID:73229121

A:Accession: A10061
 A:Molecule type: protein
 A:Residues: 27-91, 'N', 93-203, 'N', 205-266 <SHO>
 R:Shotton, D.M.; Hartley, B.S.
 Nature 225, 811-816, 1970
 A:Title: Three-dimensional structure of tosyl-elastase.
 A:Reference number: A93160; MUID:70114044
 C:Contents: annotation; X-ray crystallography, 3.5 angstroms; active site
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; pancreas; serine proteinase; zymogen
 F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-26/Domain: activation peptide #status predicted <APT>
 F:27-266/Product: elastase I #status experimental <MAT>
 F:27-259/Domain: trypsin homology <TRY>
 F:56-72,153-220,184-200,210-240/Disulfide bonds: #status experimental
 F:71,119,214/Active site: His, Asp, Ser #status experimental

Query Match 5.4%; Score 113; DB 1; Length 266;
 Best Local Similarity 27.3%; Pred. No. 0.056;
 Matches 36; Conservative 21; Mismatches 41; Indels 34; Gaps 6;

QY 147 NYPESTSVKLTG-----CTGTVAEKHVLTAHCIDHDKTY--VKGTOKLRVGFLEKPK 198
 DB 37 SWPQSLQYREGSSWMAHTCGTLLRRNMVMTAAHCVDRELFVVGEMH----- 88
 QY 199 FDDGGRGNDSTAMPQMKFQWIRKRVKTHVPGKMGINDIGMDYVALLL-KKPK 257
 DB 89 -----NNDGTE-----QYGVQKIVHPYV-----NTDVAAGYDALLRLAQSVTL 131
 QY 258 RKFMRKIGVSPRA 269
 DB 132 NSYVQLGVTPRA 143

RESULT 14

TRSMG
 trypsin (EC 3.4.21.4) precursor - Streptomyces griseus

C:Species: Streptomyces griseus
 C:Date: 24-Apr-1984 #sequence_revision 12-May-1994 #text_change 07-May-1999
 C:Accession: J01302; A00962
 R:Kim, J.C.; Cha, S.H.; Jeong, S.T.; Oh, S.K.; Byun, S.M.
 Biochem. Biophys. Res. Commun. 181, 707-713, 1991
 A:Title: Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin ge
 A:Reference number: J01302; MUID:92095977
 A:Accession: J01302

A:Molecule type: DNA
 A:Residues: 1-259 <KIM>
 A:Cross-references: GB:M64471
 A:Experimental source: strain ATCC10137
 R:Olafson, R.W.; Jurasek, L.; Carpenter, M.R.; Smillie, L.B.
 Biochemistry 14, 1168-1177, 1975
 A:Title: Amino acid sequence of Streptomyces griseus trypsin. Cyanogen bromide fragme
 A:Reference number: A00962; MUID:75127940

A:Accession: A00962
 A:Molecule type: protein
 A:Residues: 37-95, 98-259 <OLA>
 R:Read, R.J.; James, M.N.G.
 J. Mol. Biol. 200, 523, 1988
 A:Title: Refined crystal structure of Streptomyces griseus trypsin at 1.7 angstroms r
 A:Reference number: A44574; MUID:88286735
 A:Contents: annotation; X-ray crystallography, 1.7 angstroms
 A:Note: residues 96-97 modeled as Gly-Ala
 C:Genetics:

A:Gene: SPRT
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; serine proteinase
 F:1-32/Domain: signal sequence #status predicted <SIG>
 F:33-36/Domain: propeptide #status predicted <PRO>
 F:37-258/Product: trypsin #status experimental <MAT>
 F:37-252/Domain: trypsin homology <TRY>
 F:56-74,177-192,204-233/Disulfide bonds: #status experimental
 F:73,118,208/Active site: His, Asp, Ser #status experimental

Query Match 5.4%; Score 112.5; DB 1; Length 259;
 Best Local Similarity 26.0%; Pred. No. 0.06;
 Matches 39; Conservative 21; Mismatches 53; Indels 37; Gaps 5;

QY 148 YPESTSVKLTGCTGTVAEKHVLTAHCIDHDKTYVKGTKLRVGFLEKPKFDGGRGAN 207
 DB 48 FPF--VYRLSMGCGGALYADIVLTAHCV-----SSGNN 81
 QY 208 DS---TAMPQMKFQWIRKRVKTHVPGKMGINDIGMDYVALLLKKPKRKKPKIG 264

Db 82 TSIFATGCVLDQSSSAVKVSTKVLQA-----PGYNGTGKDWALIKLAQPINQPTLKIA 136
QY 265 VSPPAKOLPGGRIRHFSGYDNDPGLVYRF 294
Db 137 TTTAVNQ---GTFYVAGWGANREGSQOQRY 163

RESULT 15

E85765

hypothetical protein Z2592 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence.revision 16-Feb-2001 #text.change 31-Mar-2001

C:Accession: E85765

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,

Nature 409, 529-533, 2001.

Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

Reference number: A85480; MUID:21074935; PMID:1120551

Accession: E85765

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <STO>

A:Cross-references: GB:AE005174; NID:912315576; PIDN:AAG56585.1; GSPDB:GN00145; UWGP:Z25

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z2592

Query Match 5.48; Score 112.5; DB 2; Length 273;

Best Local Similarity 23.38; Pred. No. 0.064;

Matches 62; Conservative 35; Mismatches 112; Indels 57; Gaps 13;

QY 101 VGIYILSSSGDGAQHRDSSGSKSRKROIYGDSRFSIFGKDFLLNYPFSTSVKLTSG- 159
Db 9 LGAISLISAEVFAADKPDYAKSANDDEVSTLFEHDDRPV---NDTQSPMDAVGQLETAS 65
QY 160 ---CTGLVAEKHVLTAACIHDKTYVKGTOKLRVGLKPKFKDGR-----GANDS 209
Db 66 GNLCTATLTAAPNLALTAGHCL---LTPPKAKADKAVALRFVSNKGLMRYDIDHIEGRVDP 122
QY 210 TSAMEQOKFQWIRKRVHFKGKIKGNANDIGMDYALLEKPKHKKRKIKGVSP-- 267
Db 123 T--LGRRLKAD-----GDGWIVPPA---AAPWDFGLIVLRNPPS-----GITPLP 162
QY 268 -----PAKOLPGGRIRHFSGYDNDPGLV-YRCDVKD-ETVDLLYQOCDAPGA 315
Db 163 LFEQKALTLAKKAKGKRYQAGYFEDHDLTIYSHONCEVTGMAQTSVMASHQCDTLPGD 222
QY 316 SGSQYVVRMKRQOKWKRKIIGIFS 341
Db 223 SGSPLML---HTDQGWQ-LIGVQS 242

Search completed: July 12, 2001, 11:43:32
Job time: 197 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:42:36 ; Search time 13.12 Seconds
(without alignments)
999,989 Million cell updates/sec

Title: US-09-554-933-3
Perfect score: 2080
Sequence: 1 MAGIGLLFLFLCAVQ.....LKYAQICWIKNTLDCREG 383

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 segs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	194	9.3	316	1	GSEP_BACLI
2	135.5	6.5	313	1	MRR_BACSU
3	124	6.0	266	1	ELI_BOVIN
4	121.5	5.8	490	1	FA10_RABIT
5	119	5.7	269	1	EL2_PIG
6	115.5	5.6	269	1	EL2A_HUMAN
7	115.5	5.6	271	1	EL2_MOUSE
8	115	5.5	376	1	FA10_TROCA
9	114.5	5.5	238	1	TR15_AEDAE
10	113	5.4	266	1	ELI_PIG
11	113	5.4	266	1	ELI_RAT
12	112.5	5.4	259	1	TR15_STRGR
13	112.5	5.4	488	1	FA10_HUMAN
14	111.5	5.4	875	1	NETR_HUMAN
15	111	5.3	253	1	CEAD_HUMAN
16	111	5.3	761	1	NETR_MOUSE
17	110	5.3	492	1	EL2B_HUMAN
18	110	5.3	269	1	FA10_BOVIN
19	108	5.2	583	1	CFAI_HUMAN
20	108	5.2	786	1	ST1B_DROME
21	107.5	5.2	686	1	MAS2_HUMAN
22	105	5.0	269	1	EL2_BOVIN
23	105	5.0	274	1	TR15_ANOGA
24	104.5	5.0	268	1	TR1P_STRGA
25	103.5	5.0	271	1	EL2_RAT
26	103.5	5.0	416	1	FA9_BOVIN
27	102	4.9	902	1	ST14_MOUSE
28	101.5	4.9	281	1	TR1Z_DROER
29	98.5	4.7	260	1	COGS_HYPLI
30	98.5	4.7	267	1	TR17_ANOGA
31	97.5	4.7	452	1	FA9_CANFA
32	97	4.7	1019	1	TRC_TACTRA
33	96.5	4.6	252	1	TR1I_DROME

34	96.5	4.6	259	1	DEF3_DERPA
35	96.5	4.6	275	1	FA9_RABIT
36	96.5	4.6	454	1	TMS3_HUMAN
37	96.5	4.6	2145	1	CYAA_PODAN
38	96	4.6	274	1	TR1I_ANOGA
39	96	4.6	855	1	ST14_HUMAN
40	95.5	4.6	430	1	SNK_DROME
41	95	4.6	1019	1	LFC_CARO
42	94.5	4.5	261	1	DER3_CARO
43	93.5	4.5	227	1	TR1P_SACER
44	93.5	4.5	277	1	KLKD_HUMAN
45	93	4.5	247	1	MCT5_MOUSE

ALIGNMENTS

```

RESULT 1
GSEP_BACLI STANDARD: PRT; 316 AA.
ID GSEP_BACLI
AC P80057;
DF 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC
  ENDOPEPTIDASE) (GSE).
GN BLASE.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxID=1402;
RX MEDLINE=93054737; PubMed=1429718;
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,
  Okamoto H., Shin M., Tamaki M., Yeraoka H., Tsuzuki H., Yoshida N.;
  "Purification, characterization, cloning, and expression of a
  glutamic acid-specific protease from Bacillus licheniformis ATCC
  14580.";
RT J. Biol. Chem. 267:23782-23788(1992).
RN [2]
RP SEQUENCE OF 95-316.
RX MEDLINE=92155199; PubMed=1346764;
RA Svendsen I., Bredam K.;
  "Isolation and amino acid sequence of a glutamic acid specific
  endopeptidase from Bacillus licheniformis.";
RL Eur. J. Biochem. 204:165-171(1992).
CC -I- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE
  CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG
  PREFERENCE FOR GLU.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
  V8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D10060; BAA00949.1;
CC PIR: S23078; S23078.
CC PIR: A45134; A45134.
CC MEROPS: S01.271;
CC InterPro: IPR00126;
CC InterPro: IPR001254;
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00839; V8PROTEASE.
CC PROSITE: PS00672; V8_SER; 1.
CC PROSITE: PS00673; V8_SER; 1.

```

KW Hydrolyase; Serine protease; Signal.
 FT SIGNAL 1 2
 FT PROPER 94
 FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.
 FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 126 142
 FT DISULFID 275 279
 SQ SEQUENCE 316 AA; 33611 MW; 96D/552CB7089B09 CRC64;

Query Match 9.38; Score 194; DB 1; Length 316;
 Best Local Similarity 24.08; Pred. No. 4,4e-09;
 Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;

QY 63 SSCGPOCHKRPL---PYEEAKOYLSEETLYANGSRTEOVGYIILSSGDAQHDSG 119
 DB 28 AQAAPSPH--TPVSSDPY-KAETSVIDP-----NIKSDOYGLXSAFTGTGVNETKE 79
 QY 120 SSGSRRRKQIYGYDSRFSIFGKDFLNLN-----YPESTVYKST---GCTGLVAEKHV 170
 DB 80 KAEKSPAKKAY---SISVIGSDDRVTNTATVPYRAIVHSSISGCTGMMIGPKTV 136
 DB 171 LTAHCIDHDKT-YKGTQKLRVGLKRFKFDGGRGANDSTSNAPKPKQIMVKKRTHV 229
 DB 137 ATAGHCIDYTSSGSPAGTATVSPG-----RNGTS-----YPIGSVSTRFTI 178
 QY 230 PKGWIKNANDIGMDYDALLKPKHKRKKIGVSPAPQLPGRIFHSGYDNDRPGN 289
 DB 179 PSGMWSGNTN-----YDGAIELESPIONVYGYRYSITTSIVGTITISGPGDKTAG 233
 QY 290 IYTRCD--VKDETLDLYQCDAPGASGSGVYRMKRRQOKRERKIT---GIEGSH 343
 DB 234 TQWQSHGPIAISEYKLAQYAM-DYFGQSGSVEPQSSRNCSPCSLAVHTNGVYG- 291
 QY 344 QWVDNMGSPQDENVAVRTLPKYAICGW 372
 DB 292 -----SSYNGRTIRLKEVFDMITNW 311

RESULT 2
 MPR_BACSU STANDARD; PRT; 313 AA.
 ID MPR_BACSU STANDARD; PRT; 313 AA.
 AC P39790; 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).
 GN MPR.
 OS Bacillus subtilis.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OX Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=GP241;
 RX MEDLINE=90130256; PubMed=2105291;
 RA Sloma A., Rudolph C.F., Ruto G.A. Jr., Sullivan B.J., Theriault K.A.,
 RA Ailly D., Petro J.;
 RT "Gene encoding a novel extracellular metalloprotease in Bacillus
 RT subtilis.";
 RL J. Bacteriol. 172:1024-1029(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;
 RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
 RT Bacillus subtilis chromosome.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-68 FROM N.A.
 RX MEDLINE=89108019; PubMed=3145906;
 RX Smith H., de Jong A., Bron S., Venema G.;

RT "Characterization of signal-sequence coding regions selected from the
 RT Bacillus subtilis chromosome.";
 RT Gene 70:351-361(1988).
 RN [4]
 RP SEQUENCE OF 1-10 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96118702; PubMed=7496533;
 RA Saxild H.H., Jacobsen J.H., Nygaard P.;
 RT "Functional analysis of the Bacillus subtilis put gene encoding
 RT formate-dependent glycylamide ribonucleotide transferase.";
 RL Microbiology 141:2211-2218(1995).
 CC -1 SUBUNIT: MONOMER.
 CC -1 SUPRACELLULAR LOCATION: SECRETED.
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
 CC V8 FAMILY.
 CC -1 CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2
 CC FAMILY OF SERINE PROTEASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: L10505; AAA22604.1;
 DR EMBL: AB006424; BAA33121.1;
 DR EMBL: M22916; AAA22832.1;
 DR EMBL: Z99105; CAB12018.1;
 DR PIR: A35122; A35122.
 DR HSSP: P00756; 1SGF.
 DR MEROPS: S01.272;
 DR Subtilisin; BGI0690; mpr.
 DR Interpro: IPR000126;
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00839; V8PROTEASE.
 DR PROSITE: PS00672; V8_HIS; 1.
 DR PROSITE: PS00673; V8_SER; 1.
 KW Hydrolyase; Serine protease; Signal; Zymogen.
 FT SIGNAL 1 34
 FT PROPER 35 93
 FT CHAIN 94 313
 FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 131 147
 FT CONFLICT 61 68
 FT SEQUENCE 313 AA; 33842 MW; D41788E8D652AE94 CRC64;

Query Match 6.58; Score 135.5; DB 1; Length 313;
 Best Local Similarity 22.24; Pred. No. 0.00035;
 Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;

QY 72 GTPLPTHEAKOYLYSEETLYANGSRTEOVGYIILSSGD--GAQHRDSSGSKSRKRO 129
 DB 29 GVPKKAEMNPQTSVNTGKREDAATKNOT-----SKAQVASAPYIGTKTKTS----- 75
 QY 130 IYGYDSRF-----SIFGKD-----FLNYPSTSVKST-----GCTGL 164
 DB 76 LYGGCTELEKNIQTLQPSISITGDERRISSTTSFFPRAVQSLIKYPNTSSTYGCGFL 135
 QY 165 VAEKHVLTNAHCII-----DGKTYVGTOKLRVGLKPKFKGGRAND 208
 DB 136 VNPNTVTVAGHCYVSQDHGMASTITAAPGRNGSSYPYGY----- 175
 QY 209 STSNAPKPKQIMVKKRTHVPGKMGKN---ANDIGMDYDALLKPKHKRKKMGKIGV 265
 DB 176 -SGTFEYVSK-GWTESKPTNDYGAIKLNGSPGTWGM-YGRTTNSSP-----VGL 225
 QY 266 SPPAKQLPGGRIFHSGYDNDRPGNLVYRFDYKDETDLYQCDAPGASGSGVYRMW 325
 DB 226 SSVYVGFCDKTFGTGIMSDTPIR-----SAETKLTLY-TDVTYGCQSGSPVY----- 272

OY 326 KRQOKWKRKIIIGFSGHWDNMGSPQDENVAVRITPLKAOICW 372
Db 273 -RMSDITGOTAIHRT-----NGC-SSYNGTRVTDVFNNIQW 310

RESULT 3
ID ELL_BOVIN STANDARD; PRT; 266 AA.

AC Q28153;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).

GN ELA1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;

SEQUENCE FROM N.A.
[1]
RP STRAIN-HOUSTEIN-FRIESIAN; TISSUE-Pancreas;
RC MEDLINE-98079203; PubMed=9418008;
RA Gestin M., le Huereu-Luron I., Wicker-Planquart C., le Drean G.,
Chail J.C., Puigserver A., Guilloteau P.;
RT "Bovine pancreatic preproelastases I and II: comparison of nucleotide
and amino acid sequences and tissue specific expression.";
RL Comp. Biochem. Physiol. 118B:181-187(1997).

CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: M80838; AAA98525.1; -
HSSP: P00772; TEST.

DR MEROPS: S01.153; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS.1.
DR PROSITE: PS00135; TRYPSIN_SER.1.

KW Hydroxylase; Serine protease; Zymogen; Pancreas; Signal.
FT SIGNAL 1
FT PROPEP 17 26
FT CHAIN 27 266
FT DISULFID 56 72
FT DISULFID 153 220
FT DISULFID 184 200
FT DISULFID 210 240
FT ACT_SITE 71 71
FT ACT_SITE 119 119
FT ACT_SITE 214 214
FT CARBOHYD 87 87
FT CARBOHYD 241 241
SQ SEQUENCE 266 AA; 28518 MW; 927E29C69BAF67E7 CRC64;

FT PROPEP 17 26
FT CHAIN 27 266
FT DISULFID 56 72
FT DISULFID 153 220
FT DISULFID 184 200
FT DISULFID 210 240
FT ACT_SITE 71 71
FT ACT_SITE 119 119
FT ACT_SITE 214 214
FT CARBOHYD 87 87
FT CARBOHYD 241 241
SQ SEQUENCE 266 AA; 28518 MW; 927E29C69BAF67E7 CRC64;

Query Match 6.0%; Score 124; DB 1; Length 266;
Best Local Similarity 27.38; P-Id. No. 0.0026;
Matches 35; Conservative 25; Mismatches 38; Indels 30; Gaps 6;

OY 147 NYPESTSVKLTSG-----CTGLVAEKHVLTAACHIDGKFTYVKGQKLRVGLKPKRK 200
Db 37 SWPQSIISLYKSGSSWYHTCGTILIKQKWMVTAACHVDSQWTF-----RVILG----- 84

OY 201 DGGKANDSTSAPEQKMFQIRKRTYVPGKWKGNANDIGMDYALDEL-KPKPKRK 259
Db 85 DHNLSQNDGTE-----QYISVQKIVHPSW--NSNNVAAGYDIALVRLAQSATLNS 133

OY 260 FMKIGVSP 267
Db 134 YVQGLVLP 141

RESULT 4
ID FA10_RABIT STANDARD; PRT; 490 AA.

AC 019045;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE COAGULATION FACTOR X PRECURSOR (EC 3.4.21.6) (STUART FACTOR).

GN F10.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_TaxID=9986;

SEQUENCE FROM N.A.
[1]
RP MEDLINE-97256311; PubMed=9101642;
RA Pundurthi U.R., Anderson K.D., James H.L.;
RT "Characterization of a full-length cDNA for rabbit factor X.";
RL Thromb. Res. 85:503-514(1997).

CC -1- FUNCTION: FACTOR XA IS A VITAMIN K-DEPENDENT GLYCOPROTEIN THAT
CC CONVERTS PROTHROMBIN TO THROMBIN IN THE PRESENCE OF FACTOR VA,
CC CA++ AND PHOSPHOLIPID DURING BLOOD CLOTTING.

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-1-THR AND THEN
CC ARG-1-ILE BONDS IN PROTHROMBIN TO FORM THROMBIN.

CC -1- SUBUNIT: THE TWO CHAINS ARE FORMED FROM A SINGLE-CHAIN PRECURSOR
CC BY THE EXCISION OF TWO ARG RESIDUES AND ARE HELD TOGETHER BY 1 OR
CC MORE DISULFIDE BONDS.

CC -1- PTM: THE VITAMIN K-DEPENDENT, ENZYMAIC CARBOXYLATION OF SOME
CC GLUTAMIC ACID RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND
CC CALCIUM (BY SIMILARITY).

CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- PTM: THE ACTIVATION PEPTIDE IS CLEAVED BY FACTOR IXA (IN THE
CC INTRINSIC PATHWAY), OR BY FACTOR VIIA (IN THE EXTRINSIC PATHWAY)

CC (BY SIMILARITY).
CC -1- MISCELLANEOUS: CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO
CC ANOTHER SITE, BEYOND THE GLA DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: AF003200; AAB62542.1; -
DR InterPro: IPR000152; -
DR InterPro: IPR000294; -
DR InterPro: IPR000561; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR InterPro: IPR001881; -
DR InterPro: IPR002383; -
DR Pfam: PF00008; EGF. 2.
DR Pfam: PF00594; gla. 1.
DR Pfam: PF00089; trypsin.1.

DR InterPro: IPR000152; -
DR InterPro: IPR000294; -
DR InterPro: IPR000561; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR InterPro: IPR001881; -
DR InterPro: IPR002383; -
DR Pfam: PF00008; EGF. 2.
DR Pfam: PF00594; gla. 1.
DR Pfam: PF00089; trypsin.1.

```

DR PRINTS; PRO0001; GLABLOOD.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR PROSITE; PS00022; EGF-1; 1.
DR PROSITE; PS00186; EGF-2; 2.
DR PROSITE; PS01187; EGF-CA; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Glycoprotein; Hydroxylase; Serine protease; Plasma; blood coagulation;
KM Gamma-carboxyglutamic acid; Hydroxylation; Calcium-binding; Vitamin K;
KW Signal; Zymogen; EGF-like domain; Repeat.
FT SIGNAL; 1 20
FT PROPEP; 21 40
FT CHAIN; 41 180
FT CHAIN; 184 490
FT PROPEP; 184 232
FT CHAIN; 233 490
FT DOMAIN; 86 122
FT DOMAIN; 125 165
FT DOMAIN; 233 490
FT MOD_RES; 46 46
FT MOD_RES; 47 47
FT MOD_RES; 54 54
FT MOD_RES; 56 56
FT MOD_RES; 59 59
FT MOD_RES; 60 60
FT MOD_RES; 65 65
FT MOD_RES; 66 66
FT MOD_RES; 69 69
FT MOD_RES; 72 72
FT MOD_RES; 75 75
FT MOD_RES; 79 79
FT MOD_RES; 103 103
FT ACT_SITE; 274 274
FT ACT_SITE; 320 320
FT ACT_SITE; 417 417
FT DISULFID; 90 101
FT DISULFID; 95 110
FT DISULFID; 112 121
FT DISULFID; 129 140
FT DISULFID; 136 149
FT DISULFID; 151 164
FT DISULFID; 172 340
FT DISULFID; 239 244
FT DISULFID; 259 275
FT DISULFID; 388 402
FT DISULFID; 413 441
FT CARBOHD; 61 61
FT CARBOHD; 187 187
FT CARBOHD; 205 205
FT SEQUENCE; 450 AA; 53965 MW; 3A39FA85AF2A6D11 CMC64;

Query Match 5.8%; Score 121.5; DB 1; Length 490;
Best Local Similarity 33.3%; Pred. No. 0.0088;
Matches 32; Conservative 15; Mismatches 24; Indels 25; Gaps 4;
Oy 160 CTGTLVAEKHVLTAAHCHDGTYYKGTOKLAVGFLAKRFKDGAGANDSISAMEQMKF 219
D6 259 CGGTILSEVHYLTAAHCHDQAKRF-----KAVVGDRTDEHEBG-----NEETHEVEVYVKH 309

```

```

Oy 220 QMTRVAKTHVPGKWIKNANDIGMDYALLEKKP 255
DB 310 NRE-VKEIY-----DFDIATVLRKTP 329

RESULT 5
ID EL2_PIG STANDARD; PRT; 269 AA.
AC P08419;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 2 PRECURSOR (EC 3.4.21.71).
GN ELA2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87217962; PubMed=3646943;
RA Kawashima I., Tani T., Shimoda K., Takiguchi Y.;
RT "Characterization of pancreatic elastase II cDNAs: two elastase II
RN mRNAs are expressed in human pancreas.";
RL DNA 6:163-172(1987).
RP SEQUENCE FROM N.A.
RX MEDLINE=88198076; PubMed=2834346;
RA Shitazu Y., Yoshida H., Matsuki S., Takemura K., Ikeda N.,
RA Shimada Y., Ozawa T., Miyayama T., Iijima H., Ishida A., Sato Y.,
RA Tama I., Tanaka J., Ikenaga H.;
RT "Molecular cloning and expression in Escherichia coli of a cDNA
RT encoding human pancreatic elastase 2.";
RL J. Biochem. 102:1555-1563(1987).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: LEU-I-XAA, MET-I-XAA
CC AND PHE-I-XAA. HYDROLYSES ELASTIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; M16651; AAA31027.1; -
DR EMBL; D00237; BAA00166.1; -
DR PIR; A26823; A26823.
DR MEROPS; S01.135; -
DR InterPro; IPR001254; -
DR InterPro; IPR001314; -
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Serine protease; Pancreas; Zymogen; Signal.
FT SIGNAL; 1 16
FT PROPEP; 17 28
FT CHAIN; 29 269
FT DISULFID; 58 74
FT DISULFID; 155 222
FT DISULFID; 186 202
FT DISULFID; 212 243
FT ACT_SITE; 73 73
FT ACT_SITE; 121 121
FT ACT_SITE; 216 216
FT CONFLICT; 10 10
L -> S (IN REF. 2).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

FT CARBOHYD 186 186 N-LINKED (GLCNAc...)
SQ SEQUENCE 376 AA; 42455 MM; F5A5C7FE9BA561C1 CRC64;

Query Match
Best Local Similarity 23.8%; Pred. No. 0.022; DB 1; Length 376;
Matches 54; Conservative 29; Mismatches 72; Indels 72; Gaps 11;

66 GPOC-----HKG-----TPLLTYE--EAKOYLSTETLYANGS-----RTE 98
47 GDGCSNCHRGTCGKDGIGSTCTCLPNECKNEKVLQSCARDNGNCHMFCRQVSE 106
99 TQVGI---YIISSGSDGQOHRDSSGSKS--RRKQIYGYDSRFSIFGKDFILNYPFSTS 153
107 TQSCAESYRLGVGDHSCVAEGDFSCGRNIKARKNIKVMGDMCK-----LGECPPQAV 158
154 VKLSTG---CTGLVAEKHYLTAACHIDCKTYVKGTOKRLGFLKPKFGKGRANDST 210
159 LINEKGEVFCGGTILSPILHVLTAACHINOTKS-VKETRL-----197

211 SAMPEOMFOWIRKRVKTHVPGK--WIKGNANDIGMDYDALLELKKP 255
198 -----LSVDKRIYHTRKFPVPPNYVYVHQNPDRAVDIATIRMTTP 238

RESULT 9
TRY5_AEDAE STANDARD; PRT; 238 AA.
ID TRY5_AEDAE P29787;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE TRYPsin 561 PRECURSOR (EC 3.4.21.4) (FRAGMENT).
OS Aedes aegypti (yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Culicidae; Aedes.
OX NCBI_Taxid=7159;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97242546; PubMed=9087545;
RA Kalkok S., Tabak L.M., Prosser D.E., Downer A.E.R.,
RA White B.N.;
RT "Isolation, sequencing and characterization of two cDNA clones coding
RT for trypsin-like enzymes from the midgut of Aedes aegypti.";
RL Insect Mol. Biol. 2:71-79(1993).
-1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD
MEAL.
-1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
-1- SUBCELLULAR LOCATION: EXTRACELLULAR.
-1- TISSUE SPECIFICITY: MIDGUT.
-1- SIMILARITY: BELONGS TO TRYPSIN FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X64363; CAA45715.1; -
DR PIR: S19891; TRWV5Y.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.112;
DR InterPro: IPR001254; -
DR Pfam: PF00080; Trypsin. 1.
DR PROSITE: PS00134; TRYPsin_HIS. 1.
DR PROSITE: PS00135; TRYPsin_SER. 1.
RM Hydrolyase; Serine protease; zymogen; digestion; Multigene family.
FT NON_TER 1 1
FT PROPEP <1 11 ACTIVATION PEPTIDE.

FT CHAIN 12 238 TRYPsin 561.
FT ACT_SITE 53 53 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 97 97 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 38 54 BY SIMILARITY.
FT DISULFID 162 178 BY SIMILARITY.
FT DISULFID 189 213 BY SIMILARITY.
FT SITE 187 187 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 238 AA; 25269 MM; 3BA22FE2A32E4B5 CRC64;

Query Match
Best Local Similarity 29.4%; Pred. No. 0.014; DB 1; Length 238;
Matches 42; Conservative 22; Mismatches 44; Indels 35; Gaps 7;

149 PESTS---VKLTSGTGLVAEKHYLTAACHIDCKTYVKGTOKRLGFLKPKFGKGRG 205
24 PFOVSLGSGSSHFCGSLSERWMTAGHCAASQTLN---OVRIG--SSQHASGG-- 75
206 ANDSTSAMPEOMFOWIRKRVKTH--VPKGMINGNANDIGMDYDALLELKKPKRFRKIG 264
76 -----QLIKVKVNRHRK-----YDEVTTDYDFALLELEETVTFSDSCAP 115
265 VSPPAKQLP---GGRHFGSDYN 284
116 VKLPQKDTFVNEGTCLOVSGWGN 138

RESULT 10
EIL_PIG STANDARD; PRT; 266 AA.
ID EIL_PIG P00772; Q29625;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1986 (Rel. 07, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).
GN ELA1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_Taxid=9823;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=86304235; PubMed=3528137;
RA Shiraasu Y., Yoshida H., Miyayama T., Matsuki S., Tanaka J.,
RA Ikenaga H.;
RT "Isolation and expression in Escherichia coli of a cDNA clone
RT encoding porcine pancreatic elastase.";
RL J. Biochem. 99:1707-1712(1986).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=87250343; PubMed=3648024;
RA Tani T., Kawashima I., Furukawa H., Ohmine T., Takiguchi Y.;
RT "Characterization of a silent gene for human pancreatic elastase I:
RT structure of the 5'-flanking region.";
RL J. Biochem. 101:591-599(1987).
RN (3)
RP SEQUENCE OF 27-266.
RX MEDLINE=73229121; PubMed=4578945;
RA Shotton D.M., Hartley B.S.;
RT "Evidence for the amino acid sequence of porcine pancreatic
RT elastase.";
RL Biochem. J. 131:643-675(1973).
RN (4)
RP SEQUENCE OF 27-266.
RX MEDLINE=70114042; PubMed=5415108;
RA Shotton D.M., Hartley B.S.;
RT "Amino-acid sequence of porcine pancreatic elastase and its
RT homologues with other serine proteinases.";
RL Nature 225:802-806(1970).
RN (5)
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=78111463; PubMed=628010;


```

OY 147 NPSTSVKLTSG-----CTGTVAEKHYLTAACIHDKTY--VKGTQKRVGLPK 198
DB 37 SWPQSIHQYRSGSWAHVCGTILIRNMVMTAAHCVDRELFRRVVGHNH----- 88
OY 199 FNDGGRGANDSTSAPEQMKFQWIRYKRVKGMKGNANDIGDYALLEL-KKPK 257
DB 89 -----NNDGTE-----OYGVOKIVHPYV---NTDDVAGYDIALRLAOSVTL 131
OY 258 RKFMRIGVSPPA 269
DB 132 NSYVOLGVLPRA 143

RESULT 11
ID ELL_RAT STANDARD: PRT: 266 AA.
AC P00773:
NC 21-JUL-1986 (Rel. 01, Created)
DE 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).
GN ELA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Pictet R.L.,
RA Nikovits W., Rutter W.D.;
RT "Primary structure of two distinct rat pancreatic preproelastases
RT determined by sequence analysis of the complete cloned messenger
RT ribonucleic acid sequences."
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Lahaie R.C.,
RA Rutter W.J., McDonald R.J.;
RT "Structure of the two related elastase genes expressed in the rat
RT pancreas."
RL J. Biol. Chem. 259:14271-14278(1984).
RN [3]
RP SEQUENCE OF 17-45.
RT TISSUE-Pancreas;
RL MEDLINE=84000385; PubMed=6555050;
RA Lajman C.;
RT "Isolation and characterization of rat pancreatic elastase."
RL Biochemistry 22:3763-3770(1983).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.
CC -1- PREFERENTIAL CLEAVAGE: ALA-|-XAA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPsin FAMILY. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00116; AAA98811.1; JOINED.
DR PIR; A00960; ELRT1.
DR HSP; P00772; 1EST.
DR MEROPS; S01.153; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Pancreas; zymogen; signal.
FT SIGNAL 1 16
FT PROPEP 17 26
FT CHAIN 27 266
FT DISULFID 56 72
FT DISULFID 153 220
FT DISULFID 184 200
FT DISULFID 210 240
FT ACT_SITE 71 71
FT ACT_SITE 119 119
FT ACT_SITE 214 214
FT CONFLICT 104 104
FT CONFLICT 108 108
FT CONFLICT 244 244
FT CONFLICT 266 266
SQ SEQUENCE 266 AA; 28976 MW; 5A56FE8FCF1AAEDA CRC64;

```

```

DR EMBL; L00116; AAA98811.1; JOINED.
DR PIR; A00960; ELRT1.
DR HSP; P00772; 1EST.
DR MEROPS; S01.153; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Pancreas; zymogen; signal.
FT SIGNAL 1 16
FT PROPEP 17 26
FT CHAIN 27 266
FT DISULFID 56 72
FT DISULFID 153 220
FT DISULFID 184 200
FT DISULFID 210 240
FT ACT_SITE 71 71
FT ACT_SITE 119 119
FT ACT_SITE 214 214
FT CONFLICT 104 104
FT CONFLICT 108 108
FT CONFLICT 244 244
FT CONFLICT 266 266
SQ SEQUENCE 266 AA; 28976 MW; 5A56FE8FCF1AAEDA CRC64;

Query Match 5.4%; Score 113; DB 1; Length 266;
Best Local Similarity 26.7%; Pred. No. 0.022;
Matches 39; Conservative 23; Mismatches 48; Indels 36; Gaps 7;

OY 128 ROIVGYDSRFSIFGKDFLLNPFSTSVKLTSG-----CTGTVAEKHYLTAACIHDKT 182
DB 26 RVVGAEARBNWSPQISLQY-----LSGGSWYHTCGTILIRNMVMTAAHCVSQMT 78
OY 183 YKGTQKRVGLPKRFGDGRGANDSTSAPEQMKFQWIRYKRVKGMKGNANDIG 242
DB 79 F-----RVVVG-----DHNLSDQDGE-----OYVSVOKIVHPYV---NSNVA 115
OY 243 MDYDVALLEL-KKPKRKFMRIGVSP 267
DB 116 AGYDIALRLAOSVTLNNYVOLAVLP 141

RESULT 12
TRYP_STRGR STANDARD: PRT: 259 AA.
ID TRYP_STRGR
AC P00775;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TRYPSIN PRECURSOR (EC 3.4.21.4) (SGT).
GN SPRT.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 10137;
RL MEDLINE=92095977; PubMed=1755852;
RA Kim J.C., Cha S.H., Jeong S.T., Oh S.K., Byun S.M.;
RT "Molecular cloning and nucleotide sequence of Streptomyces griseus
RT trypsin gene."
RL Biochem. Biophys. Res. Commun. 181:707-713(1991).
RN [2]
RP SEQUENCE OF 37-259.
RX MEDLINE=75127940; PubMed=804314;
RA Olafson R.W., Jurasek L., Carpenter M.R., Smillie L.B.;
RT "Amino acid sequence of Streptomyces griseus trypsin. Cyanogen
RL bromide fragments and complete sequence."
RL Biochemistry 14:1168-1177(1975).

```

REMARKS	RESIDUE	SEQUENCE	FUNCTION
[3]	DR	EMBL: M64471; AAA26820.1; ALT_SEQ.	
DR	DR	PIR: A00962; TRSMG.	
DR	DR	PIR: J01302; J01302.	
DR	DR	PDB: 1SGT: 16-JUL-88.	
DR	DR	MEROPS: S01.101; -.	
DR	DR	InterPro: IPR001254; -.	
DR	DR	InterPro: IPR001314; -.	
DR	DR	Pfam: PF000089; trypsin, 1.	
DR	DR	PRINTS: PR00722; CHYMOTRYPSIN.	
DR	DR	PROSITE: PS00134; TRYPsin_HIS; 1.	
DR	DR	PROSITE: PS00135; TRYPsin_SER; 1.	
KW	KW	Hydrolase; Serine protease; zymogen; Signal; 3D-structure.	
FT	FT	SIGNAL	1 32
FT	FT	PROPEP	33 36
FT	FT	CHAIN	37 239
FT	FT	ACT_SITE	73 73
FT	FT	ACT_SITE	118 118
FT	FT	ACT_SITE	208 208
FT	FT	DISULFID	58 74
FT	FT	DISULFID	177 192
FT	FT	DISULFID	204 233
FT	FT	SITE	202 202
FT	FT	CONFLICT	95 96
FT	FT	TURN	38 38
FT	FT	TURN	39 39
FT	FT	TURN	41 42
FT	FT	TURN	45 46
FT	FT	TURN	49 50
FT	FT	TURN	51 54
FT	FT	TURN	55 57
FT	FT	TURN	58 64
FT	FT	TURN	65 66
FT	FT	TURN	67 70
FT	FT	HELIX	72 74
FT	FT	STRAND	79 80
FT	FT	STRAND	85 88
FT	FT	STRAND	92 92
FT	FT	TURN	93 94
FT	FT	TURN	96 97
FT	FT	STRAND	99 108
FT	FT	TURN	110 111
FT	FT	STRAND	120 124
FT	FT	STRAND	134 135
FT	FT	TURN	140 141
FT	FT	STRAND	145 150
FT	FT	TURN	156 157
FT	FT	STRAND	163 163
FT	FT	STRAND	165 172
FT	FT	HELIX	174 184
FT	FT	TURN	187 189
FT	FT	STRAND	190 193
FT	FT	TURN	196 198
FT	FT	TURN	202 202
FT	FT	STRAND	205 206

	Query Match	Best Local Similarity	5.4%	Score 112.5;	DB 1;	Length 259;	
	Matches 39;	Conservative	21;	Mismatches 53;	Indels 37;	Gaps 5;	
OY	148 YPFSTVSKLSTCTGTVAEHKHLTAACIHDKTIVKGOKLRVGFLPKPFRDGGGAN	207					
	: : :	:		:	:	:	
Db	48 PFF-MWRLSMCGGALVADQIYLTAHCY-----SGSGNN	81					
OY	208 DS---TSAMEQMKFOWIRKTHVPKCWGKNANDIGMDYYALLEKKPHKRKFMMKG	264					
	- - - -:- - -:- :: : :	:: :	:	:	:	:	
Db	82 TSITATGVGDVDLOSSAAVKVRSTKVLA----PGYNGTKDMALKIAQTNPPTLKIA	136					
OY	265 VSPPAKQLPCGRIFHSGYDNDRGNLVYRF	294					
	- - - -:- - -:- :: : :	:: :	:	:	:	:	
Db	137 TTTAVNQ--GTFTVAGMGANEGGSQQRY	163					
	RESULT 13						
	FA10_HUMAN	STANDARD:	PRT:	488 AA.			
AC	P00742;						
DT	21-JUL-1986 (Rel. 01, Created)						
DT	01-OCT-1989 (Rel. 12, Last sequence update)						
DT	01-OCT-2000 (Rel. 40, Last annotation update)						
DE	COAGULATION FACTOR X PRECURSOR (EC 3.4.21.6) (STUART FACTOR).						
F10.							
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
OX	NCB1-TaxID=9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=91216473; PubMed=1902434;						
RA	Messier T.L., Pittman D.D., Long G.L., Kaufman R.J., Church W.R.;						
RT	"Cloning and expression in COS-1 cells of a full-length cDNA encoding						
RL	human coagulation factor X.";						
	Gene 99:291-294(1991).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=87026600; PubMed=3768336;						
RA	Leytus S.P., Foster D.C., Kurachi K., Davie E.W.;						
RT	"Gene for human factor X: a blood coagulation factor whose gene						
RL	organization is essentially identical with that of factor IX and						
	protein C.";						
RT	Biochemistry 25:5098-5102(1986).						
RN	[3]						
RP	SEQUENCE OF 13-488 FROM N.A.						
RX	MEDLINP=85216545; PubMed=2582420;						
RA	Fung M.R., Hay C.W., McGillyvray R.T.A.;						
RT	"Characterization of an almost full-length cDNA coding for human						
RL	blood coagulation factor X.";						
	Proc. Natl. Acad. Sci. U.S.A. 82:3591-3595(1985).						
RN	[4]						
RP	SEQUENCE OF 19-488 FROM N.A.						
RC	TISSUE=Liver;						
RX	MEDLINE=86221713; PubMed=3011603;						
RA	Kaul R.K., Hildebrand B., Roberts S., Jagadeeswaran P.;						
RT	"Isolation and characterization of human blood-coagulation factor X						
RL	cDNA.";						
	Gene 41:311-314(1986).						
RN	[5]						
RP	SEQUENCE OF 41-179.						

DR	PIR: A25853; A25853.	
DR	PIR: A24478; A24478.	
DR	PDB: 1HCG; 08-MAY-95.	
DR	PDB: 1FAK; 29-OCT-97.	
DR	PDB: 1EXY; 17-JUN-98.	
DR	PDB: 1XKA; 23-MAR-99.	
DR	PDB: 1XKB; 23-MAR-99.	
DR	MEROPS: S01.216; -.	
DR	GLYCOSULEDB: P00742; -.	
DR	CarbBank; CCSD:29393; -.	
DR	MIM: 134530; -.	
DR	MIM: 227600; -.	
DR	InterPro: IPR000152; -.	
DR	InterPro: IPR000294; -.	
DR	InterPro: IPR000561; -.	
DR	InterPro: IPR001254; -.	
DR	InterPro: IPR001314; -.	
DR	InterPro: IPR001881; -.	
DR	InterPro: IPR002383; -.	
DR	Pfam: PF00008; EGF; 1.	
DR	Pfam: PF00594; g1a; 1.	
DR	Pfam: PF00089; trypsin; 1.	
DR	PRINTS: PR00001; GLABLOOD.	
DR	PRINTS: PR00722; CHYMOTRYPSIN.	
DR	PROSITE: PS00010; ASX_HYDROXYL; 1.	
DR	PROSITE: PS00011; GLU_CARBOXYLATION; 1.	
DR	PROSITE: PS00022; EGF_1; 1.	
DR	PROSITE: PS00134; TRYPSIN_HIS; 1.	
DR	PROSITE: PS00135; TRYPSIN_SER; 1.	
DR	PROSITE: PS01186; EGF_2; 2.	
DR	PROSITE: PS01187; EGF_CA; 1.	
KW	Gamma-carboxyglutamyl acid; Hydroxylase; Serine protease; Plasma; Blood coagulation; Calcium-binding; Vitamin K	
KW	Signal; Zymogen; EGF-like domain; Repeat; 3D-structure.	
KW	SIGNAL	1
FT	PROPEP	1
FT	CHAIN	40
FT	CHAIN	179
FT	CHAIN	183
FT	PROPEP	183
FT	CHAIN	234
FT	CHAIN	235
FT	DOMAIN	86
FT	DOMAIN	122
FT	DOMAIN	125
FT	DOMAIN	165
FT	DOMAIN	235
FT	MOD_RES	46
FT	MOD_RES	46
FT	MOD_RES	47
FT	MOD_RES	47
FT	MOD_RES	54
FT	MOD_RES	54
FT	MOD_RES	56
FT	MOD_RES	56
FT	MOD_RES	59
FT	MOD_RES	59
FT	MOD_RES	60
FT	MOD_RES	60
FT	MOD_RES	65
FT	MOD_RES	65
FT	MOD_RES	66
FT	MOD_RES	66
FT	MOD_RES	69
FT	MOD_RES	72
FT	MOD_RES	72
FT	MOD_RES	79
FT	MOD_RES	79
FT	MOD_RES	103
FT	MOD_RES	103
FT	CARBOHYD	199
FT	CARBOHYD	199
FT	CARBOHYD	211
FT	CARBOHYD	221
FT	CARBOHYD	231
FT	CARBOHYD	231
FT	ACT_SITE	276
FT	ACT_SITE	276
FT	ACT_SITE	322
FT	ACT_SITE	419
FT	DISULFID	90
FT	DISULFID	101
FT	DISULFID	110
FT	DISULFID	110
FT	DISULFID	112
FT	DISULFID	121
FT	DISULFID	129
FT	DISULFID	140
FT	DISULFID	136
FT	DISULFID	149
FT	DISULFID	151
FT	DISULFID	164
FT	DISULFID	172
FT	DISULFID	342
FT	DISULFID	241
FT	DISULFID	246
FT	DISULFID	261
FT	DISULFID	277
FT	DISULFID	390
FT	DISULFID	404
FT	DISULFID	415
FT	DISULFID	443

FT CONFLICT 285 288 KVRV -> E (IN REF. 4 AND 6).
 FT CONFLICT 442 442 G -> S (IN REF. 3).
 SQ SEQUENCE 488 AA; 54731 MW; F81D5746AF4797AF CRC64;

Query Match
 Best Local Similarity 24.4%; Score 111.5; DB 1; Length 488;
 Matches 49; Conservative 28; Mismatches 59; Indels 63; Gaps 10;

DB 160 CTGTLVAEKHVLPAACIHGKTYVKTOKLRGFLKPKKDGGRANDSTAMPDQMF 219
 261 CGGTILSEFILLPAHCLYAKRF---KVRGDRNTGEGEGEAVHE-VEVYIKHNR 314
 220 QWIKRTHYHPKGIKGNANDIGMDYALLETLEKPKRKFPMKIGVSPPA----- 269
 315 -----TKETY-----DFDIALVRLKTPIT---FRNNVAPACLPERDMAEST 352
 QY 270 --KOLPG---GRHESGYDNDPRPNLYVRFCDVKD-----ETYDLLYQ 307
 DB 353 LMTQKTGIVSGFGRTHKEGRQSTRLEKMLEVYVDNRNSCKLSSEFILTQNNFCAGYDT--K 410
 308 QCDAPGASGSGVYVPMK 326
 411 QEDACQGDGSG-GRHVTYRK 428

RESULT 14
 NETR_HUMAN STANDARD; PRT; 875 AA.

AC 36730; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEUROTRYPsin PRECURSOR (EC 3.4.21.-) (MOTOPsin).
 GN PRSS12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98201705; PubMed=9540828;
 RA Proba K., Geschwend T.P., Sonderegger P.;
 RT "Cloning and sequencing of the cDNA encoding human neurotropsin.";
 RL Biochim. Biophys. Acta 1396:143-147(1998).
 CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
 CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
 CC LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 SRCR DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL: A001531; GAA04816.1; -
 DR InterPro: IPR000001; -
 DR InterPro: IPR001190; -
 DR InterPro: IPR001254; -
 DR InterPro: IPR001314; -
 DR PRINTS: PRO0258; SPERACTRCPR.
 DR PRINTS: PRO0722; CHYMOTRYPsin.
 DR PROSITE: PS00134; TRYPSIN_HTS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.

KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 875
 FT DOMAIN 23 92
 FT DOMAIN 93 165
 FT DOMAIN 170 271
 FT DOMAIN 280 381
 FT DOMAIN 387 487
 FT DOMAIN 500 601
 FT DOMAIN 619 875
 FT DOMAIN 619 875
 FT ACT SITE 630 631
 FT ACT SITE 676 676
 FT ACT SITE 726 726
 FT ACT SITE 825 825
 FT DISULFID 619 750
 FT CARBOHYD 26 26
 FT CARBOHYD 683 683
 SQ SEQUENCE 875 AA; 97011 MW; B66EC946DC208DC8 CRC64;

Query Match
 Best Local Similarity 23.1%; Score 111.5; DB 1; Length 875;
 Matches 73; Conservative 47; Mismatches 85; Indels 111; Gaps 21;

QY 118 SGSSGK-----RRKQIYGYDSRFSIGKDEL-NYPFSTVKLSTG----- 159
 DB 607 SGNSKKESSVSGGLRLRRQKRIIG-----GKNSLRGCMPOVSLRLKSSHGDR 658
 QY 160 --CTGTLVAEKHVLPAACIHGKTYVKTOK--LRVGLRKFKFDGGANDSTAMPE 215
 DB 659 ILGCAVLLSSCWLLPAHCLYAKRF---KRYGNSTRSYAVRAG-----DYTLVPE 701
 QY 216 QMK---FQWIRYKRVHVPKGIKGNANDIGMDYALLETLEKPKR--KFMKIGVSPPA 269
 DB 702 EFEELGVQQIVYHREYRD-----RSYDIALVRLGQPEQCARFSS-HYLPAC 750
 QY 270 -----KOLPGRIHPSGY-DNDPRGN-----LYVRFCDVKD---TYDLL--- 305
 DB 751 LPLMRERPKTASNCYITGWDGTRAVSRTLQOMALPLLPKRCERKRYGRTGRMLCAG 810
 QY 306 -----YQCCDAPGASGSGVYVPMKROQKWEKKIIGFSHOWVMNGSPQDFNVAARI 361
 DB 811 NLEHKKVDSQGDSDGSL---MCEPGESEW-VYIGVTS---W-----GYGGVAD 854
 QY 362 TPLKYAQICY---WIK 374
 DB 855 SPGYITVSAFVPMK 870

RESULT 15
 CFAD_HUMAN STANDARD; PRT; 253 AA.

AC P00746;
 DT 21-JUN-1986 (Rel. 01, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COMPLEMENT FACTOR D PRECURSOR (EC 3.4.21.46) (C3 CONVERTASE ACTIVATOR)
 DE (PROPERDIN FACTOR D) (ADIPSIN).
 GN DF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 8-253 FROM N.A.
 RX MEDLINE=92250520; PubMed=1374388;
 RA White R.T., Damm D., Hancock N., Rosen B.S., Lowell B.B., Usher P.,
 RA Flier J.S., Spiegelman B.M.;
 RT "Human adiponin is identical to complement factor D and is expressed
 at high levels in adipose tissue.";
 RL J. Biol. Chem. 267:9210-9213(1992).
 RN [2]

RP SEQUENCE FROM N.A.
RA Flier J.S., Spiegelman B.M., Rosen B.M.;
RL Patent number WO9006365, 14-JUN-1990.
RN [3]
RP SEQUENCE OF 26-252.
RX MEDLINE-85000441; PubMed-6383466;
RA Niekman M.A., Bhown A.S., Bennett J.C., Volanakis J.E.;
RT "Amino acid sequence of human D of the alternative complement
pathway."
RL Biochemistry 23:2482-2486(1984).
RN [4]
RP PARTIAL SEQUENCE OF 26-252.
RX MEDLINE-84108950; PubMed-6363133;
RA Johnson D.M.A., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence of human factor D of the complement system.
Similarity in sequence between factor D and proteases of non-plasma
origin."
RL FEBS Lett. 166:347-351(1984).
RN [5]
RP PARTIAL SEQUENCE OF 26-61 AND 194-220.
RX MEDLINE-84256515; PubMed-6821372;
RA Johnson D.M.A., Gagnon J., Reid K.B.M.;
RT "Factor D of the alternative pathway of human complement.
Purification, alignment and N-terminal amino acid sequences of the
major cyanogen bromide fragments, and localization of the serine
residue at the active site."
RL Biochem. J. 187:863-874(1980).
RN [6]
RP PARTIAL SEQUENCE OF 26-82.
RX MEDLINE-80145719; PubMed-6987665;
RA Volanakis J.E., Bhown A.S., Bennett J.C., Mole J.E.;
RT "Partial amino acid sequence of human factor D: homology with serine
proteases."
RL Proc. Natl. Acad. Sci. U.S.A. 77:1116-1119(1980).
RN [7]
RP PARTIAL SEQUENCE OF 26-78.
RX MEDLINE-81054886; PubMed-6776531;
RA Davis A.E. III;
RT "Active site amino acid sequence of human factor D."
RL Proc. Natl. Acad. Sci. U.S.A. 77:4938-4942(1980).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-94118317; PubMed-8289289;
RA Narayana S.V.L., Carson M., El-Kabbani O., Kilpatrick J.M., Moore D.,
Chen X., Bugb C.E., Volanakis J.E., Delucas L.J.;
RT "Structure of human factor D. A complement system protein at 2.0-A
resolution."
RL J. Mol. Biol. 235:695-708(1994).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-96025834; PubMed-7592653;
RA Kim S., Narayana S.V., Volanakis J.E.;
RT "Crystal structure of a complement factor D mutant expressing
enhanced catalytic activity."
RL J. Biol. Chem. 270:24399-24405(1995).
RN [10]
RP FUNCTION: FACTOR D CLEAVES FACTOR B WHEN THE LATTER IS COMPLEXED
WITH FACTOR C3B, ACTIVATING THE C3BB COMPLEX, WHICH THEN BECOMES
THE C3 CONVERTASE OF THE ALTERNATE PATHWAY. ITS FUNCTION IS
HOMOLOGOUS TO THAT OF C1S IN THE CLASSICAL PATHWAY.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY.
CC - CAUTION: IN ADDITION TO THE CONFLICTS SHOWN IN THE FEATURE TABLE,
REF.3 SEQUENCE HAD A PEPTIDE IN THE WRONG ORDER AND ANOTHER ONE
MISSING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL; M84526; AAA35527.1; ALT_INT.
DR PIR; A40197; DBHU.
DR PDB; 1DFP; 25-FEB-98.
DR PDB; 1DST; 11-JUL-96.
DR PDB; 1DSU; 17-AUG-96.
DR MEROPS; S01.191; -.
DR MIM; 134350; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR Complement alternate pathway; Plasma; Hydrolyase; Serine protease;
KW Zymogen; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 20
FT PROPEP 1 25
FT CHAIN 26 253
FT ACT_SITE 66 66
FT ACT_SITE 112 112
FT ACT_SITE 208 208
FT DISULFID 51 67
FT DISULFID 148 214
FT DISULFID 179 195
FT DISULFID 204 229
FT CONFLICT 26 26
FT CONFLICT 35 35
FT CONFLICT 40 40
FT CONFLICT 49 49
FT CONFLICT 52 52
FT CONFLICT 59 59
FT CONFLICT 63 63
FT CONFLICT 73 73
FT CONFLICT 83 86
FT CONFLICT 84 84
FT CONFLICT 93 95
FT CONFLICT 96 96
FT CONFLICT 136 136
FT CONFLICT 178 191
FT CONFLICT 243 243
FT CONFLICT 250 250
FT CONFLICT 250 250
FT SEQUENCE 253 AA; 27004 MW; BD53B70BD5C6AD CRC64;
SO
Query Match 5.3%; Score 111; DB 1; Length 253;
Best Local Similarity 28.4%; Pred. No. 0.03; Mismatches 43; Indels 48; Gaps 9;
Matches 44; Conservative 20;
QY 149 PFSTSVKLSTG--CGTLVAKRHVLTAAHCIIH--DGKTYVKGTOKLKRVGFLKPKFKDGG 203
DB 38 PYMASVQNLGAHLGGVLAEGQWVLSAAHCLDADAGVQVL----- 79
QY 204 RGANDSTNMEPQMKFQWIRKRVKTHVPRKGIKGNANDIGMDYVALLLELKKPKRKPKTKI 263
DB 80 LGAHLSLSPPEESKRYLDVLAFA-----VP-----HPDQPTIDHLLDLQLSE-----KA 124
QY 264 GVSPPAKQLPGGRHIFSGYDND-RPGNVYRCDV 297
DB 125 TLGPAVRPLPMQRV-----DRDVAFGTL-----CDV 150

Search completed: July 12, 2001, 11:44:23
Job time: 107 sec

Thu Jul 12 15:25:01 2001

us-09-554-933-3.rsp

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:41:55 ; Search time 24.68 seconds
(without alignments)
2053.194 Million cell updates/sec

Title: US-09-554-933-3
Perfect score: 2080
Sequence: 1 MAGIPGLFLFLCAVQ.....LKYAQICWIKGNLDCREG 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_proteus:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

.Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2080	100.0	383	4 095084	095084 homo sapien
2	129.5	6.2	303	2 09EXR9	09EXR9 bacillus in
3	129	6.2	469	6 09GMD9	09GMD9 ornithorhyn
4	127.5	6.1	1322	5 09NATO	09NATO anophelies g
5	126.5	6.1	678	11 09JUS8	09JUS8 rattus norv
6	124.5	6.0	1322	5 09NUS5	09NUS5 anophelies g
7	124	6.0	266	6 046644	046644 macaca fasc
8	123.5	5.9	339	11 09QX91	09QX91 rattus norv
9	123.5	5.9	366	11 09QX85	09QX85 rattus norv
10	123.5	5.9	541	11 09QX90	09QX90 rattus norv
11	123.5	5.9	623	11 09JUP3	09JUP3 rattus norv
12	123.5	5.9	643	11 09QX84	09QX84 rattus norv
13	122	5.9	259	5 09X161	09X161 ctenocephal
14	122	5.9	449	5 09VDO8	09VDO8 drosophila
15	120.5	5.8	482	11 063207	063207 rattus norv
16	118	5.7	1376	5 09VOR8	09VOR8 drosophila
17	117	5.6	258	5 09W508	09W508 drosophila
18	117	5.6	522	5 001771	001771 caenorhabdi
19	115.5	5.6	269	4 014243	014243 homo sapien

20	115	5.5	258	4 09UN11	09UN11 homo sapien
21	113.5	5.5	418	5 09VA87	09VA87 drosophila
22	113.5	5.5	685	11 092338	092338 mus musculu
23	113.5	5.5	737	13 090422	090422 brachydanto
24	113	5.4	573	5 09V516	09V516 drosophila
25	112.5	5.4	273	2 09V176	09V176 escherichia
26	112.5	5.4	603	3 09VDV1	09VDV1 drosophila
27	112	5.4	249	13 09W7Q1	09W7Q1 paralicthy
28	110.5	5.3	279	5 096991	096991 manduca sex
29	110.5	5.3	481	11 054740	054740 mus musculu
30	110	5.3	1449	5 010922	010922 caenorhabdi
31	109.5	5.3	405	2 069973	069973 streptomyce
32	108	5.2	274	5 017086	017086 anophelies s
33	108	5.2	750	13 09W633	09W633 cyprinus ca
34	108	5.2	1047	5 024019	024019 drosophila
35	107.5	5.2	236	11 0921H1	0921H1 mus musculu
36	107.5	5.2	481	11 088947	088947 mus musculu
37	107.5	5.2	611	5 09VZ58	09VZ58 drosophila
38	107	5.1	603	11 061129	061129 mus musculu
39	106	5.1	460	5 09VFP6	09VFP6 drosophila
40	106	5.1	762	13 09VIC6	09VIC6 cyprinus ca
41	105.5	5.1	268	13 09W7Q2	09W7Q2 paralicthy
42	105.5	5.1	604	11 09WUM3	09WUM3 rattus norv
43	105	5.0	787	5 09VER6	09VER6 drosophila
44	104.5	5.0	290	5 09VRT2	09VRT2 drosophila
45	104	5.0	1190	5 09V7B5	09V7B5 drosophila

ALIGNMENTS

RESULT 1
ID 095084 PRELIMINARY; PRT; 383 AA.
AC 095084:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN).
GN ZSIG13 OR DKFP586B0719.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UMBILICAL VEIN;
RA Li X., Tedder T.F.,
RT "A novel serine protease from human umbilical vein endothelial
RT cells.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TISSUE=UTERUS;
RC SEQUENCE FROM N.A.
RA Mambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL: AF015287; ADO01553.1; -;
DR EMBL: AF193611; AAF07186.1; -;
DR EMBL: AL136914; CAB66848.1; -;
DR MEROPS; S01.309; -;
DR InterPro; IPR001254; -;
DR InterPro; IPR001314; -;
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR SMART; SM00020; TRYP_SPC; 1.
KW Hydrolase; Protease; Serine protease.

SEQUENCE 383 AA; 43001 MW; 46EB6C11ABFDE8F CRC64;

Query Match 100.0%; Score 2080; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 2e-181;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLCAVGVSPYSAFWKPTWAVRLPVLPSTLNLAKPDGAEAKLE 60
DB 1 MAGIPGLFLFLFLCAVGVSPYSAFWKPTWAVRLPVLPSTLNLAKPDGAEAKLE 60
QY 61 VSSCGQCHKGTPLEPTYEBAKQYLSYETLYANGSRTETGVGIIYLLSSGDAQHDSGS 120
DB 61 VSSCGQCHKGTPLEPTYEBAKQYLSYETLYANGSRTETGVGIIYLLSSGDAQHDSGS 120
QY 121 SGKSRRRKQIYGYDSRFSIFGKDFLNPFPSTSVKLSGCTGLVAKHVLTAHCHDQ 180
DB 121 SGKSRRRKQIYGYDSRFSIFGKDFLNPFPSTSVKLSGCTGLVAKHVLTAHCHDQ 180
QY 181 KTYVKGTOGLRVGFLKPKFKDGRGANDSTSAMPEQMKFQWIRVKTTHVFKWIKGNAND 240
DB 181 KTYVKGTOGLRVGFLKPKFKDGRGANDSTSAMPEQMKFQWIRVKTTHVFKWIKGNAND 240
QY 241 IGMDDYVALLLEKPKPKFKKFKKIGVSPAKQLPGRRIHFSGYDNDPGRNLYRRCQVKE 300
DB 241 IGMDDYVALLLEKPKPKFKKFKKIGVSPAKQLPGRRIHFSGYDNDPGRNLYRRCQVKE 300
QY 301 TYDLYQOCDAOPGASGGVYVYRMKROQKWKERKIIGIFSGHGVDMNGSPQDENVAVR 360
DB 301 TYDLYQOCDAOPGASGGVYVYRMKROQKWKERKIIGIFSGHGVDMNGSPQDENVAVR 360
QY 361 ITPKYAOICWIKGNVLDRCRG 383
DB 361 ITPKYAOICWIKGNVLDRCRG 383

RESULT 2
Q9EXR9 PRELIMINARY; PRT: 303 AA.
AC Q9EXR9; 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
OS GLUTAMYL-ENDOPEPTIDASE.
OC Bacillus intermedius.
OC Bacillus/Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1400;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99170141; PubMed=10071925;
RABRIKOV D.V., AKIMKINA T.V., SHEVELEV A.B., DEMIDYUK I.V.,
BUSHNEVA A., KOSTOV S.V., CHESTUKHINA G.G., STEPANOV V.M.;
"Bacillus intermedius glutamyl endopeptidase. Molecular cloning and
RT nucleotide sequence of the structural gene.";
RT J. Protein Chem. 18:21-25(1999).
DR EMBL: Y15136; CAC17594.1;
SQ SEQUENCE 303 AA; 32343 MW; F2E37B5B07781D59 CRC64;

Query Match 6.2%; Score 129.5; DB 2; Length 303;
Best Local Similarity 24.6%; Pred. No. 0.00088;
Matches 65; Conservative 40; Mismatches 98; Indels 61; Gaps 15;

QY 108 SSGGAGHRDSSGSGSRKROIYGYDSRFSIFGKDFLNPFPSTSVKLSGTG---CTGL 164
DB 68 SSTSGEVYKPLSKYLKFOFKYVIGDDGKRVANTRA---PINSIAYITFGSGCTGL 124
QY 165 VAEKHVLTAAHCHD---GKTY-VKGTOKLRYGLKPKFGKDGANDSTSAMPEQMKFQW 221
DB 125 IAPKKIILNGHCYVNTASRSYSAKGS-----VYP-----GMDSTAVNGS----- 164
QY 222 IRVKTTHVFKWIKGNANDIGMDYVALLLEKPKPKFKKIGVSPAKQLPQ 275

DB 165 ANMTEFYVPSGYINTGAS-----QTFDAVITD-----TNINQTYGNSIRQVNTLGTG 213
QY 276 RIHESGYDND---RPGNLVYRF---CDVKEDEYDLYQOCDAOPGASGGVYVYRMKROQ 330
DB 214 TIKISGYGDAMRSTGKVKSGWEMSGSVTRBEDTNLAYTYIDTFSGNSGSA---MLDONO 269
QY 331 KWERTIIGIF-SGHQWDMNGSPQ 353
DB 270 -----IVGVHNGVSNCTINGPK 288

RESULT 3
Q9GMD9 PRELIMINARY; PRT: 469 AA.
AC Q9GMD9; 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
OS COAGULATION FACTOR X.
OC Ornithorhynchus anatinus (Duckbill platypus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
OX NCBI_Taxid=9258;
RN [1]
RP SEQUENCE FROM N.A.
RA Poorafshar M.M., Hellman L.L.;
RT "Identification and structural analysis of three serine proteases in a
RT monotreme, platypus, Ornithorhynchus anatinus.";
RT Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF275654; AAC00453.1;
SQ SEQUENCE 469 AA; 52196 MW; 4C66C23D00758F6A CRC64;

Query Match 6.2%; Score 129; DB 6; Length 469;
Best Local Similarity 26.5%; Pred. No. 0.0018;
Matches 66; Conservative 30; Mismatches 73; Indels 80; Gaps 16;

QY 160 CTGTLVAKKHVLTAAHCHDGTYYKGTOKLRYGLKPKFKDGANDSTSAMPEQMKF 219
DB 258 CGGTILNEYIILSAACHMOAKRF-----KVRGGERDTERKKSSEMAHEVERIVHS-KF 311
QY 220 QMIRVKTTHVFKWIKGNANDIGMDYVALLLEKPKPKFKKIGVSP---PAK----- 270
DB 312 ---VKRTY-----DEDIAYIKIKTPTIT---FRMNVSPACLPKEDMAEDI 349
QY 271 ---QLPG-----GRHESGYDNDPQNL-----YRFDVKEDEYDLYQO-C----- 309
DB 350 LMNQKAGVYSGFGRVHEKG---RPSVILKMLEVYRPTCKROSSSFDITPMMFCAGTD 405
QY 310 ----DAQPGASGGVYVYRMKROQKWKERKIIGIFSGHGVDMNGSPQDENVAVRITPLK 365
DB 406 SRPDACQGDGSG-GPHVTKYKDTY-----FVIGIVMGECCAGNGK---FGVYT-----K 451
QY 366 YAOICWYWK 374
DB 452 AATFLSWIK 460

RESULT 4
Q9NATO PRELIMINARY; PRT: 1322 AA.
AC Q9NATO; 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
OS ADHESIVE SERINE PROTEASE.
SP22D.
GN Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anophelinae.
OX NCBI_Taxid=7165;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE-20318993; PubMed-10860981;
RA Dantelli A., Loukeris T., Lagneux M., Mueller H.M., Richman A.,
Katates F.C.;
RT "A modular chitin-binding protease associated with hemocytes and
hemolymph in the mosquito *Anopheles gambiae*.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:7136-7141(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(LDLRA) DOMAIN.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL, AJ276428; CAB81934.1; -.
DR InterPro: IPR001190; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR002172; -.
DR InterPro: IPR002557; -.
DR Pfam: PF00057; 1dl_recept_a; 2.
DR Pfam: PF00089; trypsin; 1.
DR Pfam: PF00530; SRCR; 2.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRPT.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 2.
DR PROSITE: PS00420; SRCR_1; UNKNOWN_1.
DR PROSITE: PS0287; SRCR_2; 2.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR SMART: SM00494; Chtd2; 1.
DR GlycoProtein: Protease.
KW SEQUENCE 1322 AA; 146794 MW; FB973C21CC5475B CRC64;

Query Match 6.1%; Score 127.5; DB 5; Length 1322;
Best Local Similarity 25.3%; Pred. No. 0.0097;
Matches 46; Conservative 35; Mismatches 58; Indels 43; Gaps 9;

OY 128 ROIVYDGRSIFGKDFLLNPESTSVKLT--GCTGLVNAKHYLTAAHCHDKTYK 185
DB 1078 RYVHSE---TVYG-----HHWQASLRKTKMHCAGAVLITRYHYLTAAHCLIG--YPK 1126
OY 186 GTOKLRVGLFKPKFKDGRGANDSTSAMPEQMKFQWIRKTHVPGWIKGNANDIGDY 245
DB 1127 STYRRLG-----DYHTAAYDNAELD-IFENYTHIQREGH-----HMSN 1167
OY 246 DYALIELKKPKRKKRWKIGVSPAKQ---LPGRIHSGYDNDPRGNLVYRCDVDETY 302
DB 1168 DIAVAVLTKPVRFNQYVOPICLPARDAPYLPQONCTISGWGTEAGS-----KDSY 1219
OY 303 DL 304
DB 1220 DL 1221
RESULT 5
O9JUS8 PRELIMINARY; PRT; 678 AA.
AC O9JUS8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE MANNOSE-BINDING PROTEIN ASSOCIATED SERINE PROTEASE-2 PRECURSOR
DE (FRAGMENT).
GN MASP-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Wallis R., Dodd R.B.;

RT "Interaction of mannose-binding protein with associated serine
RT proteases: Effects of naturally occurring mutations.";
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL, AJ277747; CAB90832.1; -.
DR InterPro: IPR000152; -.
DR InterPro: IPR000436; -.
DR InterPro: IPR000561; -.
DR InterPro: IPR000859; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR001881; -.
DR Pfam: PF00084; sushl; 2.
DR Pfam: PF00089; trypsin; 1.
DR Pfam: PF00431; CUB; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR SMART: SM00032; CCP; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat;
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 12 POTENTIAL.
FT CHAIN 13 678 MANNOSE-BINDING PROTEIN ASSOCIATED SERINE
FT PROTEASE-2.
SQ SEQUENCE 678 AA; 74734 MW; 16680E4A9ADCC05 CRC64;

Query Match 6.1%; Score 126.5; DB 11; Length 678;
Best Local Similarity 22.1%; Pred. No. 0.0049;
Matches 79; Conservative 46; Mismatches 123; Indels 109; Gaps 19;

OY 65 GCP-----QCHKPPLPYEBKQYLYETLYANGSRTEOVGIYILSSGSGAQH 115
DB 359 CQPPDPLNGHVDTYITGEVYTKAVIQYSCHEFTYTMSSN-----GRYVEADGFWTSS 413
OY 116 RDS-----GSSGSRKRRQIYGYDSRSISGKFLNYPSTSVKLTGCTGLV 165
DB 414 KGEKLPVCKPQYGLSTHTSGRIITGGQPAKG-----DEPQVILLGETTAAGALI 465
OY 166 AEKHYLTAAHCHDKTYVYKTKLRVGLFKPKFKDGRGANDSTSAMPEQMKFQWIRK 225
DB 466 HDQWVLTAAHANY-KTEAMSSLDTRMGILK-----RLSHYQANPEAV----- 509
OY 226 RTHVPGWIKGNANDIGMDYDALLELK--PHKRRKPKIGVSPAKQLPG-----GR 276
DB 510 --FIHEGYTHG---AGPDNDIALIKLKNTYINENIMPICL--PRKEAASIMKTDVFGT 561
OY 277 IHFSGYDNDPRG---NLVYRCDVYKDEYDLYOQC-----DAQPGA----- 315
DB 562 V--AGWGLTKQKFLARNLMFVDIPVD-----HQCATAVTKQPYPAKYTVNNMLCAGL 613
OY 316 -----SGSGVYVRMKRROOKKERRKIIGIFSGHGVMDN--GSPQDFNVAVRIT 362
DB 614 DAGKDSGRGDSGALVFLDNETQRM--FVGSIYS---WGSINGSGSRYGYTVTVT 665
RESULT 6
O9NUS5 PRELIMINARY; PRT; 1322 AA.
AC O9NUS5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SERINE PROTEASE 22D.
GN SP22D.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.

Db 324 CGPDDLPNGHVDYITGTEVYTKAVIOYSCFEFFMYTSSN-----GRYCEADGEFTSS 378
Oy 116 RDS-----GSSGSRKROIYGYDSRPSFGKDFLLNTPSTSVKLTGCTFLV 165
Db 379 KGEKSLPVCYKPGCGSTSTSGRIIGGOPARPG-----DFPQVILLGETTAGALI 430
Oy 166 AEKHLVTAHNCIHDKTYVKTOKLVGFLKPKFGKRGANDSTSAPEQKQWIRVK 225
Db 431 HDDWVLTAAHAYV-GKTEAMSSLDIRMGILK-----RLSLITYQAMPEAV----- 474
Oy 226 RTHPKKIKGNANDIGNDYDYLLELK--PHKRFKIKGVSPPAKOLPG-----GR 276
Db 475 --FHEGTYHG---AGPDNDIALIKLNKYITNNIPICL--PRKASASLMKTDYVGT 526
Oy 277 IHFSGYDNDPRG---NLYRRCQVCKDEYDLYQOC-----DAQPGA----- 315
Db 527 V--AGWGLTQGFARNLNMFDPIDV-----HOKCATATKQYPPAKATYVNLGAGL 578
316 -----SSGYYVYRMMKROQOQKERRKIIGFSGHQMVMN-GSPDDFNVAVRIT 362
Db 579 DRGKDCRGDSGALVFLDNETGRW--FVGIGVS---WGSINCGSGBOGYVTKVT 630

RESULT 13
Oy 09xy61 PRELIMINARY; PRT; 259 AA.
AC 09xy61;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE TRIPSN-LIKE SERINE PROTEASE (FRAGMENT).
GN SP-6.
OS Ctenocephalides felis (Cat flea).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Siphonaptera; Pulicidae;
OC Ctenocephalides.
OC NCBI_TaxID=7515;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99124360; PubMed=9927170;
RA Gaines P.J., Sampson C.M., Rushlow K.E., Stieglitz G.L.;
RT *Cloning of a family of serine protease genes from the cat flea
Ctenocephalides felis.*
RL Insect Mol. Biol. 81:11-22(1999).
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
EMBL: AF053919; AAD1839.1; -
HSSP: P00763; IDPO.
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR Pfam: PFO0089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR SMART: SM00020; TRYP-Ser; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
SO SEQUENCE 259 AA; 28299 MW; 98222E7EB430472A CRC64;

Query Match 5.98; Score 122; DB 5; Length 259;
Best Local Similarity 24.18; Pred. No. 0.0034;
Matches 49; Conservative 30; Mismatches 68; Indels 56; Gaps 10;

Oy 101 VGIYILSS--SGDQNHDSGSSGKRRKQIYGYDSRPSIFGKDFLLNTPSTSVKLT 158
Db 11 VGLSASVYKTKKDGIRIVGQADAIK---YGTQASLQVFNHF----- 53
Oy 159 GCTGLVAEKHLVTAHNCIHDKTYVKTOKLVGFLKPKFGKRGANDSTSAPEQMK 218
Db 54 -CGASILNNWIVTAHNCIHDYDEFY-----SVRVG---TSFQ-GARGSVHVAQIILKIPA 103
Oy 219 FQWIRVKTHTVPKKIKGNANDIGNDYDYLLELKP--HKRKPWKIGVSPAKOLPG 275

Db 104 Y-----GNVTDI--DMEXALIKVRRPRLNRTVTRVTKLDVGDMPG 145
Oy 276 RHFF-----SGYDNDPRGNLY 292
Db 146 ELATVTGMNIGDEDDPEQLOY 168

RESULT 14
Oy 09yD08 PRELIMINARY; PRT; 449 AA.
AC 09yD08;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE CG7432 PROTEIN.
GN CG7432.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
EMBL: AE003727; AAE55692.2; -
CC FLYbase: FBgn0038727; CG7432.
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR Pfam: PFO0089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.

us-09-554-933-3.rsp

Page 8

Query Match	5.9%;	Score 122;	DB 5;	Length 443;
Best Local Similarity	23.6%;	Pred. No. 0.0072;		
Matches	62;	Conservative	81;	Indels 88; Gaps 13
			Mismatches	

[illegible]

RESULT	15		
ID	063207	PRELIMINARY;	PRT; 482 AA.
AC	063207;		
DT	01-NOV-1996 (TREMBLREL. 01,	Created)	
DT	01-NOV-1996 (TREMBLREL. 01,	Last sequence update)	
DT	01-MAR-2001 (TREMBLREL. 16,	Last annotation update)	
DE	FACTOR X.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SPRAGUE-DAWLEY;		
FX	MEDLINE=66093366; Pubmed=8578539;		
RA	Stanton C., Ross R.P., Hutson S., Wallin R.;		
RT	"Evidence for competition between vitamin K-dependent clotting factors		
RT	for intracellular processing by the vitamin K-dependent gamma-		
RT	carboxylase";		
RT	Thromb. Res. 80:63-73(1995).		
CC	-1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.		
CC	-1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).		
EMBL	X79807: CAA56202.1; -		
DR	HSSP; P00742; 1XKA.		
DR	MEROPS; S01216; -		
DR	InterPro; IPR000152; -		
DR	InterPro; IPR000294; -		
DR	InterPro; IPR0000561; -		
DR	InterPro; IPR000742; -		
DR	InterPro; IPR001254; -		
DR	InterPro; IPR001314; -		
DR	InterPro; IPR001881; -		
DR	InterPro; IPR002383; -		
DR	PIfam; PF00008; EGF_2.		
DR	PIfam; PF00089; trypsin; 1.		
DR	PIfam; PF00594; gla; 1.		
DR	PRINTS: PR00722; CHYMOTRYPSIN.		
DR	PRINTS: PR00001; GLABLOD.		
DR	PROSITE: PS00010; ASX_HYDROXYL_1.		
DR	PROSITE: PS00022; EGF_1; UNKNOWN_1.		
DR	PROSITE: PS01186; EGF_2; 2.		
DR	PROSITE: PS01187; EGF_CA_1.		

Query Match	5.88;	Score 120.5;	DB 11;	Length 482;	
Best Local Similarity	32.38;	Pred. No. 0.011;			
Matches	31;	Conservative	17;	Mismatches	23;
				Indels	25;
				Gaps	4

QY	160	CTGTLVAEKHVLTAALNCCHIDGKTYKQKQKLRGELPKRKDGGRANASVLRK	311
		: : : : : : : : : : :	
Db	259	CGGILMEFIILPAALCLQAKRF-----KRVADLTFTBEDGEGENVHE-VMIITKHKFF	314
QY	220	QMIRKRTAHVKGKMGKGNANDIGMDYDALLLEKRP	255
		: : : : : : : : : : : : : : : : :	
Db	313	Q-----RDLY-----DFDIAMLRLKTP	329

Search completed: July 12, 2001, 11:44:05
Job time: 130 sec

THIS PAGE BLANK (USPTO)